STIC-Biotech/ChemLib 73796

From: Chan, Christina
Sent: Tuesday, August 20, 2002 3:19 PM
To: Wilder, Cynthia; STIC-Biotech/ChemLib
Subject: RE: Rush sequence search 09/835,082

12E11

Please rush. Thanks Chris

-----Original Message-----
From: Wilder, Cynthia
Sent: Tuesday, August 20, 2002 12:14 PM
To: Chan, Christina
Subject: Rush sequence search 09/835,082

Ms. Chan,

I am requesting a Rush sequence search for 09/835,082 due to a low docket. Please forward your approval to STIC.

Please search the nucleic acid encoded by SEQ ID NO: 1 and SEQ ID NO: 3.

Thank you!

Cynthia B. Wilder, Ph.D.
Art Unit 1637
703/305-1680
Office No. 12E11
Mailbox No. 10E12

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 8/20/02
Date Completed: 8/21/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: 2
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: 05
WWW/Internet: _____
Other (specify): _____

```
Qy   1252 gttctggaaatgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaa 1311
          || || ||||||||| |||| ||||| |||||||| || ||| | ||||||||  ||||
Db    421 gtccttgaaatgactgctgatgatttggaaaatgcccttaaaactgtcatcaacaacaag 480

Qy   1312 agttacaaggagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccg 1371
          || ||||||||||||||||||||||||||||||||||||||||||| ||  | |||||
Db    481 agctacaaggagaacatcatgcgcctctccagccttcacaaggaccgtcctatagagcct 540

Qy   1372 ctggacctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctg 1431
          |||||||||| |||||||||||||| |  |||||||||||||||| || |||||||||
Db    541 ctggacctggctgtgttctgggtggaatacgtgatgaggcacaaggggcaccacacctg 600

Qy   1432 cgccccgcagcccacgacctcacctggtaccagtaccattccttggacgtgattggtttc 1491
          ||||| || ||||| |||||||||||||| |||||||| |||||||| |||||||| |||
Db    601 cgccggccgcccatgacctcacctggtatcagtaccactccttggatgtgattggcttc 660

Qy   1492 ctcttggccgtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctac 1551
          ||| ||||| | ||| |||||||| ||||||   |||||||||||||| ||||||| |
Db    661 ctcctggccattgtgttgacagtggtcttcattgnctttaaatgttgtgcctatggctgc 720

Qy   1552 cggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattga 1611
          |||||||||| |||  ||| |||| ||| ||||||| | ||||||||||||||||||||
Db    721 cggaaatgctttgggggaaaggggcnagtgaagaaatcacacaaatccaagacccattga 780

Qy   1612 gaagtgggtgggaaataaggtaaaattttgaaccattccct 1652
          |||||||| || |   || |  || | ||    ||||  ||
Db    781 gaagtgggggaagtgaangagaagtattagttcattatct 821
```

Search completed: August 21, 2002, 12:45:22
Job time: 22240 sec

```
CC   animals or modified cells e.g. for pharmacogenetic screening.
XX
SQ   Sequence 759 BP; 190 A; 165 C; 169 G; 235 T; 0 other;


  Query Match             28.4%;  Score 659.8;  DB 21;  Length 759;
  Best Local Similarity   91.8%;  Pred. No. 4.4e-177;
  Matches  697;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy    115 atggatgggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggg 174
          |||||||||||| ||||||||||||||  |||||||||||||||||| || |||||||||
Db      1 atggatgggagtcactggttcaccatgcagtcggtggtggagaaacttatcctcaggggg 60

Qy    175 catgaggtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgc 234
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     61 catgaggtggttgtagtcatgccagaggtgagttggcaactggaaagatcactgaattgc 120

Qy    235 acagtgaagacttattcaacttcatataccctggaggatctggaccgggagttcaaggct 294
          |||||||||||||| ||||| || || || |||| |||| | ||||||| |||| || |
Db    121 acagtgaagacttactcaacctcgtacactctggaagatcagaaccgggaattcatggtt 180

Qy    295 tttgcccatgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttca 354
          || |||||||||||||||||||||| | || |||||||| |||||||||||||| |||||
Db    181 ttcgcccatgctcaatggaaagcacaggcacaaagtatattttctctattaatgagttca 240

Qy    355 tacaatgacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaa 414
          | || ||   || ||||||||||||||||  |||||||||||||||||||| |||  ||||
Db    241 tccagtggttttcttgacttatttttttcgcattgcaggagtttgtttaatgaccgaaaa 300

Qy    415 ttagtagaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataac 474
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||| |
Db    301 ttagtagaatacttaaaggagagttcttttgatgcagtgtttctggatccttttgatacc 360

Qy    475 tgtggcttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaata 534
          |||||||||||||||| |||||||||||||||||| ||||||||| ||||||||||||
Db    361 tgtggcttaattgttgctaaatatttctccctcccctctgtggtcttcaccaggggaata 420

Qy    535 ctttgccactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccaga 594
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ttttgccaccatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccaat 480

Qy    595 attctcttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatg 654
          ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||| ||
Db    481 gatctcttagggttctcagatgccatgactttcaaggagagagtatggaaccacatcgtg 540

Qy    655 cacttggaggaacatttattatgccaccgtttttttcaaaaatgccctagaaatagcctct 714
          |||||||||| |||||||| |||||    | |||| | ||||||||||||||||||||||
Db    541 cacttggaggaccatttattttgccagtatctttttagaaatgccctagaaatagcctct 600

Qy    715 gaaattctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttg 774
          |||||||||||||| ||||| ||||  |||||||||||||| ||||||||||||||||||
Db    601 gaaattctccaaacccctgtcacggcatatgatctctacagtcacacatcaatttggttg 660

Qy    775 ttgcgaacggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcatt 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ttgcgaacggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcatt 720

Qy    835 ggtggtatcaactgccatcagggaaagccgttgcctatg 873
          |||||||||||||| |||||||||||| |||||||||
Db    721 ggtggtatcaactgtcatcagggaaagccattgcctatg 759


RESULT  15
ABL01468
ID   ABL01468 standard; DNA; 874 BP.
XX
AC   ABL01468;
XX
DT   15-MAR-2002  (first entry)
XX
DE   Murine apoptosis related DNA sequence #133.
XX
KW   Apoptosis; mouse; cancer; autoimmune disease; viral infection;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW   reperfusion injury; stroke; liver damage; dilatory cardiomyopathy;
KW   transgenic animal; hepatotropic; antialcoholism; cytostatic;
KW   immunosuppressive; virucide; nootropic; neuroprotective; vasotropic;
KW   antiparkinsonian; cerebroprotective; ds.
XX
OS   Mus sp.
XX
PN   DE10126344-A1.
XX
PD   24-JAN-2002.
XX
PF   30-MAY-2001; 2001DE-1026344.
XX
PR   14-JUL-2000; 2000DE-1034303.
XX
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Grimm S,  Schoenfeld N,  Braziulis E,  Cramer U,  Gewies A,  Voss F;
PI   Mund T,  Albayrak T,  Gille H,  Klein M;
XX
DR   WPI; 2002-115563/16.
XX
PT   New apoptosis-associated nucleic acid sequences and polypeptides,
PT   useful for diagnosis, treatment and prevention of e.g. tumors and
PT   neurodegeneration -
XX
PS   Claim 1; Page 184; 227pp; German.
XX
CC   The present invention relates to nucleic acids from the mouse, where the
CC   nucleic acid is associated with apoptosis. The sequences can be used in
CC   the diagnosis, treatment and prevention of diseases associated with
CC   excessive or inadequate apoptosis, including tumours, autoimmune
CC   diseases, viral infections, degradative diseases (Alzheimer's,
CC   Parkinson's and Huntington's diseases), reperfusion injury, stroke and
CC   alcohol-induced injury to the liver, for identifying agents for treating
CC   these diseases, and to prepare transgenic animals in which expression of
CC   an apoptosis related sequence is altered. These are useful for genetic
CC   and/or pharmacological investigations of apoptosis and related diseases,
CC   including dilatory cardiomyopathy. The present sequence is one of the
CC   apoptosis related sequences of the invention.
XX
SQ   Sequence 874 BP; 222 A; 202 C; 233 G; 210 T; 7 other;


  Query Match             24.9%;  Score 578.8;  DB 24;  Length 874;
  Best Local Similarity   84.9%;  Pred. No. 5.3e-154;
  Matches  646;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy    892 aatgcttctggagaacatggaattgtggttttctctttgggatcaatggtctcagaaatt 951
          || || ||||| || ||||| || |||||||||||||||||||| |||||||||| |||
Db     61 aacgcctctggggagcatggcatcgtggttttctctttgggatccatggtctcagagatt 120

Qy    952 ccagagaagaaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctg 1011
          || |||||||||| |||| ||||||||| |||||||||| ||| |||||||| ||||||
Db    121 ccggagaagaaagccatggaaattgctgaggctttgggcagaattcctcagacggtcctg 180

Qy   1012 tggcggtacactggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtgg 1071
          ||||| ||||| |||||  |||||||||||||||| || ||||| || ||||| || |||
Db    181 tggcgctacaccggaactagaccatcgaatcttgcaaagaacacaattcttgtcaaatgg 240

Qy   1072 ctaccccaaaacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggt 1131
          |||||||||| |||||||||||||| || | ||| || || || ||||| ||  ||||
Db    241 ctaccccaaaatgatctgcttggtcatccaaagactcgggcattcatcacacactctggc 300

Qy   1132 tcccatggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgttt 1191
          ||||||||| |||||||| | |||||||||| ||||| ||||||||||||||| | |||
Db    301 tcccatggtatttatgaaggaatatgcaatggagttccgatggtgatgatgcccctattt 360

Qy   1192 ggtgatcagatggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaat 1251
          || |||||||||||||||| |||||||||| |||  ||||||||| ||||||||||||
Db    361 ggcgatcagatggacaatgccaagcgcatggaaactcggggagctggggtgaccctgaat 420
```

```
XX
DR   WPI; 2000-052981/04.
DR   P-PSDB; AAY57100.
XX
PT   New nucleic acid representing polymorphisms in the human uridine
PT   diphosphate glucuronosyltransferase gene, used for diagnosis and
PT   evaluation of drug metabolism  -
XX
PS   Examples, Page 43-44; 63pp; English.
XX
CC   Nucleotide sequences AAZ45110-Z45118 are exons 1A-1J of the human
CC   uridine diphosphate-glucuronosyltransferase 1 (UGT1) gene. The UGTs are a
CC   family of enzymes that catalyse the glucuronic acid conjugation of a wide
CC   range of endogenous and exogenous substrates including  phenols,
CC   alcohols, amines and fatty acids. Many of the reactions catalysed by UGTs
CC   result in toxic substances being converted to compounds which are more
CC   water soluble and are excreted. The invention relates to and identifies
CC   UGT1 polymorphisms (AAZ45004-Z45041). The polymorphism sequences are
CC   useful as probes for detecting UGT1 locus polymorphisms, indicative of
CC   altered UGT1 expression or activity. These polymorphisms are associated
CC   with Crigler-Najjar and Gilbert syndromes (unconjugated
CC   hyperbilirubinaemia) and drug metabolism. The genotyping of the UGT1 gene
CC   is used to predict the rate of metabolism of UGT1 substrates, possible
CC   drug-drug interactions and adverse side effects (i.e. to optimize drug
CC   dosage), and to screen for diseases caused by exposure to toxins and to
CC   study the effects of polymorphisms on enzymatic activity. The UGT1
CC   sequences, including polymorphisms, can also be used to produce the
CC   corresponding protein (or its fragments) or to generate transgenic
CC   animals or modified cells e.g. for pharmacogenetic screening.
XX
SQ   Sequence 735 BP; 189 A; 179 C; 187 G; 180 T; 0 other;
```

```
  Query Match             31.7%;  Score 735;  DB 21;  Length 735;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-198;
  Matches  735;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     874 gaatttgaagcctacattaatgcttctggagaacatggaattgtggttttctctttggga 933
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gaatttgaagcctacattaatgcttctggagaacatggaattgtggttttctctttggga 60

Qy     934 tcaatggtctcagaaattccagagaagaaagctatggcaattgctgatgctttgggcaaa 993
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 tcaatggtctcagaaattccagagaagaaagctatggcaattgctgatgctttgggcaaa 120

Qy     994 atccctcagacagtcctgtggcggtacactggaacccgaccatcgaatcttgcgaacaac 1053
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 atccctcagacagtcctgtggcggtacactggaacccgaccatcgaatcttgcgaacaac 180

Qy    1054 acgatacttgttaagtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcc 1113
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 acgatacttgttaagtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcc 240

Qy    1114 tttatcacccatgctggttcccatggtgtttatgaaagcatatgcaatggcgttcccatg 1173
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 tttatcacccatgctggttcccatggtgtttatgaaagcatatgcaatggcgttcccatg 300

Qy    1174 gtgatgatgcccttgtttggtgatcagatggacaatgcaaagcgcatggagactaaggga 1233
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 gtgatgatgcccttgtttggtgatcagatggacaatgcaaagcgcatggagactaaggga 360

Qy    1234 gctggagtgaccctgaatgttctggaaatgacttctgaagatttagaaaatgctctaaaa 1293
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 gctggagtgaccctgaatgttctggaaatgacttctgaagatttagaaaatgctctaaaa 420

Qy    1294 gcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctccagccttcacaag 1353
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 gcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctccagccttcacaag 480

Qy    1354 gaccgcccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgaggcac 1413
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 gaccgcccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgaggcac 540

Qy    1414 aagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccattcc 1473
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 aagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccattcc 600

Qy    1474 ttggacgtgattggtttcctcttggccgtcgtgctgacagtggccttcatcacctttaaa 1533
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ttggacgtgattggtttcctcttggccgtcgtgctgacagtggccttcatcacctttaaa 660

Qy    1534 tgttgtgcttatggctaccggaaatgcttggggaaaaaagggcgagttaagaaagcccac 1593
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 tgttgtgcttatggctaccggaaatgcttggggaaaaaagggcgagttaagaaagcccac 720

Qy    1594 aaatccaagacccat 1608
           ||||||||||||||||
Db     721 aaatccaagacccat 735
```

```
RESULT  14
AAZ45117
ID   AAZ45117 standard; DNA; 759 BP.
XX
AC   AAZ45117;
XX
DT   28-FEB-2000  (first entry)
XX
DE   UDP-glucuronosyltransferase 1 (UGT1) exon 1J nucleotide sequence.
XX
KW   Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
KW   glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
KW   unconjugated hyperbilirubinaemia; drug metabolism; transgenic animal;
KW   pharmacogenetic screening; diagnose; ss.
XX
OS   Homo sapiens.
XX
PN   WO9957322-A2.
XX
PD   11-NOV-1999.
XX
PF   04-MAY-1999;   99WO-US09702.
XX
PR   07-MAY-1998;   98US-0084807.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Penny L,  Galvin M;
XX
DR   WPI; 2000-052981/04.
DR   P-PSDB; AAY57099.
XX
PT   New nucleic acid representing polymorphisms in the human uridine
PT   diphosphate glucuronosyltransferase gene, used for diagnosis and
PT   evaluation of drug metabolism  -
XX
PS   Examples; Page 42; 63pp; English.
XX
CC   Nucleotide sequences AAZ45110-Z45118 are exons 1A-1J of the human
CC   uridine diphosphate-glucuronosyltransferase 1 (UGT1) gene. The UGTs are a
CC   family of enzymes that catalyse the glucuronic acid conjugation of a wide
CC   range of endogenous and exogenous substrates including  phenols,
CC   alcohols, amines and fatty acids. Many of the reactions catalysed by UGTs
CC   result in toxic substances being converted to compounds which are more
CC   water soluble and are excreted. The invention relates to and identifies
CC   UGT1 polymorphisms (AAZ45004-Z45041). The polymorphism sequences are
CC   useful as probes for detecting UGT1 locus polymorphisms, indicative of
CC   altered UGT1 expression or activity. These polymorphisms are associated
CC   with Crigler-Najjar and Gilbert syndromes (unconjugated
CC   hyperbilirubinaemia) and drug metabolism. The genotyping of the UGT1 gene
CC   is used to predict the rate of metabolism of UGT1 substrates, possible
CC   drug-drug interactions and adverse side effects (i.e. to optimize drug
CC   dosage), and to screen for diseases caused by exposure to toxins and to
CC   study the effects of polymorphisms on enzymatic activity. The UGT1
CC   sequences, including polymorphisms, can also be used to produce the
CC   corresponding protein (or its fragments) or to generate transgenic
```

DE UDP-glucuronosyltransferase 1 (UGT1) exon 1G nucleotide sequence.
XX
KW Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
KW glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
KW unconjugated hyperbilirubinaemia; drug metabolism; transgenic animal;
KW pharmacogenetic screening; diagnose; ss.
XX
OS Homo sapiens.
XX
PN WO9957322-A2.
XX
PD 11-NOV-1999.
XX
PF 04-MAY-1999; 99WO-US09702.
XX
PR 07-MAY-1998; 98US-0084807.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Penny L, Galvin M;
XX
DR WPI; 2000-052981/04.
DR P-PSDB; AAY57097.
XX
PT New nucleic acid representing polymorphisms in the human uridine
PT diphosphate glucuronosyltransferase gene, used for diagnosis and
PT evaluation of drug metabolism -
XX
PS Examples; Page 37-38; 63pp; English.
XX
CC Nucleotide sequences AAZ45110-Z45118 are exons 1A-1J of the human
CC uridine diphosphate-glucuronosyltransferase 1 (UGT1) gene. The UGTs are a
CC family of enzymes that catalyse the glucuronic acid conjugation of a wide
CC range of endogenous and exogenous substrates including phenols,
CC alcohols, amines and fatty acids. Many of the reactions catalysed by UGTs
CC result in toxic substances being converted to compounds which are more
CC water soluble and are excreted. The invention relates to and identifies
CC UGT1 polymorphisms (AAZ45004-Z45041). The polymorphism sequences are
CC useful as probes for detecting UGT1 locus polymorphisms, indicative of
CC altered UGT1 expression or activity. These polymorphisms are associated
CC with Crigler-Najjar and Gilbert syndromes (unconjugated
CC hyperbilirubinaemia) and drug metabolism. The genotyping of the UGT1 gene
CC is used to predict the rate of metabolism of UGT1 substrates, possible
CC drug-drug interactions and adverse side effects (i.e. to optimize drug
CC dosage), and to screen for diseases caused by exposure to toxins and to
CC study the effects of polymorphisms on enzymatic activity. The UGT1
CC sequences, including polymorphisms, can also be used to produce the
CC corresponding protein (or its fragments) or to generate transgenic
CC animals or modified cells e.g. for pharmacogenetic screening.
XX
SQ Sequence 951 BP; 229 A; 211 C; 214 G; 297 T; 0 other;

Query Match 31.9%; Score 741.2; DB 21; Length 951;
Best Local Similarity 90.5%; Pred. No. 3.5e-200;
Matches 791; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```
Qy      19 atggcttgcacagggtggaccagccccctcctctatgtgtgtgtctgctgctgacctgt 78
           ||||||  |  ||||||||||   ||| ||||||  |||| |||||||||  ||||||||||
Db       1 atggctcgtgcagggtggactggcctccttcccctatatgtgtgtctactgctgacctgt 60

Qy      79 ggctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcacc 138
           |  ||  |   | ||||||||||  ||||||||||||||||||||||||||||||||||
Db      61 gctttgccaaggtcagggaagctgctggtagtgcccatggatgggagccactggttcacc 120

Qy     139 atgaggtcggtggtggagaaactcattctcagggggcatgaggtggttgtagtcatgcca 198
           |||  ||||||||||||||||||| ||||||||||||||||| ||||||||||||
Db     121 atgcagtcggtggtggagaaactcatcctcagggggcatgaggtggtcgtagtcatgcca 180

Qy     199 gaggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttca 258
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||| |||
Db     181 gaggtgagttggcaactgggaagatcactgaattgcacagtgaagacttactcaacctca 240

Qy     259 tataccctggaggatctggaccgggagttcaaggcttttgcccatgctcaatggaaagca 318
           || || |||||||||| |||||||||||||| || ||||||| ||||||  ||||  |||
Db     241 tacactctggaggatcaggaccgggagttcatggtttttgccgatgctcgctggacggca 300

Qy     319 caagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttattt 378
           | | | ||||||  || ||||||||||   |||||| |||||  ||||||||||||||||
Db     301 ccattgcgaagtgcattttctctattaacaagttcatccaatggtatttttgacttattt 360

Qy     379 ttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagagt 438
           |||||||||||||||||||||||||| |||  |||||||||||||||||||||||||||
Db     361 ttttcaaattgcaggagtttgtttaatgaccgaaaattagtagaatacttaaaggagagt 420

Qy     439 tcttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaatat 498
           | ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db     421 tgttttgatgcagtgtttctcgatccttttgatcgctgtggcttaattgttgccaaatat 480

Qy     499 ttctccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaaggt 558
           |||||||||||||| |||||||||||||||||||| |||||||||||||||||||||||
Db     481 ttctccctcccctctgtggtcttcgccaggggaatattttgccactatcttgaagaaggt 540

Qy     559 gcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgcc 618
           |||||||||||||||||||||||||||||||||||| |||||||||||||||||||  |||
Db     541 gcacagtgccctgctcctctttcctatgtccccagacttctcttagggttctcagacgcc 600

Qy     619 atgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatgc 678
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||  |||
Db     601 atgactttcaaggagagagtatggaaccacatcatgcacttggaggaacatttattttgc 660

Qy     679 caccgtttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttacg 738
           | |  |||||||||||||| | ||||||||||||||||||||||||||| ||||| |||
Db     661 ccctatttttttcaaaaatgtcttagaaatagcctctgaaattctccaaacccctgtcacg 720

Qy     739 gagtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttggac 798
           |  ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     721 gcatatgatctctacagccacacatcaatttggttgttgcgaactgactttgttttggag 780

Qy     799 tatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggga 858
           |||||||||||||||||||||||| ||||||||||||||||||||||||| |||||||||
Db     781 tatcccaaacccgtgatgcccaatatgatcttcattggtggtatcaactgtcatcaggga 840

Qy     859 aagccgttgcctatggaatttgaagcctacatta 892
           |||||  |||||||||  |  | |  |  | |||
Db     841 aagccagtgcctatggtaagttatctccccttta 874
```

RESULT 13
AAZ45118
ID AAZ45118 standard; DNA; 735 BP.
XX
AC AAZ45118;
XX
DT 28-FEB-2000 (first entry)
XX
DE UDP-glucuronosyltransferase 1 (UGT1) exons 2-5 nucleotide sequence.
XX
KW Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
KW glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
KW unconjugated hyperbilirubinaemia; drug metabolism; transgenic animal;
KW pharmacogenetic screening; diagnose; ss.
XX
OS Homo sapiens.
XX
PN WO9957322-A2.
XX
PD 11-NOV-1999.
XX
PF 04-MAY-1999; 99WO-US09702.
XX
PR 07-MAY-1998; 98US-0084807.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Penny L, Galvin M;

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 catatgatttcttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatt 660

Qy   2211 tgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatataca 2270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 tgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatataca 720

Qy   2271 gctatgaagtgctgggcaagtttacttttttttctgatgtttcctacaact 2320
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 gctatgaagtgctgggcaagtttacttttttttctgatgtttcctacaact 770


RESULT  11
AAZ45116
ID   AAZ45116 standard; DNA; 930 BP.
XX
AC   AAZ45116;
XX
DT   28-FEB-2000  (first entry)
XX
DE   UDP-glucuronosyltransferase 1 (UGT1) exon 1H nucleotide sequence.
XX
KW   Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
KW   glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
KW   unconjugated hyperbilirubinaemia; drug metabolism; transgenic animal;
KW   pharmacogenetic screening; diagnose; ss.
XX
OS   Homo sapiens.
XX
PN   WO9957322-A2.
XX
PD   11-NOV-1999.
XX
PF   04-MAY-1999;   99WO-US09702.
XX
PR   07-MAY-1998;   98US-0084807.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Penny L,  Galvin M;
XX
DR   WPI; 2000-052981/04.
DR   P-PSDB; AAY57098.
XX
PT   New nucleic acid representing polymorphisms in the human uridine
PT   diphosphate glucuronosyltransferase gene, used for diagnosis and
PT   evaluation of drug metabolism  -
XX
PS   Examples; Page 39-41; 63pp; English.
XX
CC   Nucleotide sequences AAZ45110-Z45118 are exons 1A-1J of the human
CC   uridine diphosphate-glucuronosyltransferase 1 (UGT1) gene. The UGTs are a
CC   family of enzymes that catalyse the glucuronic acid conjugation of a wide
CC   range of endogenous and exogenous substrates including  phenols,
CC   alcohols, amines and fatty acids. Many of the reactions catalysed by UGTs
CC   result in toxic substances being converted to compounds which are more
CC   water soluble and are excreted. The invention relates to and identifies
CC   UGT1 polymorphisms (AAZ45004-Z45041). The polymorphism sequences are
CC   useful as probes for detecting UGT1 locus polymorphisms, indicative of
CC   altered UGT1 expression or activity. These polymorphisms are associated
CC   with Crigler-Najjar and Gilbert syndromes (unconjugated
CC   hyperbilirubinaemia) and drug metabolism. The genotyping of the UGT1 gene
CC   is used to predict the rate of metabolism of UGT1 substrates, possible
CC   drug-drug interactions and adverse side effects (i.e. to optimize drug
CC   dosage), and to screen for diseases caused by exposure to toxins and to
CC   study the effects of polymorphisms on enzymatic activity. The UGT1
CC   sequences, including polymorphisms, can also be used to produce the
CC   corresponding protein (or its fragments) or to generate transgenic
CC   animals or modified cells e.g. for pharmacogenetic screening.
XX
SQ   Sequence 930 BP; 228 A; 205 C; 212 G; 285 T; 0 other;

  Query Match           33.0%;  Score 766.6;  DB 21;  Length 930;
  Best Local Similarity   91.6%;  Pred. No. 2.1e-207;
  Matches  811;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     19 atggcttgcacagggtggaccagccccttcctctatgtgtgtgtctgctgctgacctgt 78
          |||||| ||||||||||||||||||||| |||| |||||||| | ||||||||||||||||
Db      1 atggctcgcacagggtggaccagccccattcccctatgtgtttctctgctgctgacctgt 60

Qy     79 ggctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcacc 138
          |||||||| |||||||||||||| |||||||||||||||||||||| ||||||||||||
Db     61 ggctttgctgaggcagggaagctgctggtagtgcccatggatgggagtcactggttcacc 120

Qy    139 atgaggtcggtggtggagaaactcattctcaggggggcatgaggtggttgtagtcatgcca 198
          |||  |||||||||||||||||| || ||||||||||||||||||||||||||||||||
Db    121 atgcagtcggtggtggagaaacttatcctcaggggggcatgaggtggttgtagtcatgcca 180

Qy    199 gaggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttca 258
          |||||||||||||||||||| |||||||||||||||||||||||| ||||| |||
Db    181 gaggtgagttggcaactgggaaaatcactgaattgcacagtgaagacttactcaacctca 240

Qy    259 tataccctggaggatctggaccgggagttcaaggcttttgcccatgctcaatggaaagca 318
          || || ||||||||||||||||||||| |||| || ||| ||| ||||||||||||||||
Db    241 tacactctggaggatctggaccgggaattcatggatttcgccgatgctcaatggaaagca 300

Qy    319 caagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttattt 378
          |||||||||||| | | |||||||||  || |||||| |||||   ||||| ||||||||
Db    301 caagtacgaagtttgttttctctatttctgagttcatccaatggtttttttaacttattt 360

Qy    379 ttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagagt 438
          |||||  |||||||||||||||||| |||  ||||||||||||||||||||||||||||
Db    361 ttttcgcattgcaggagtttgtttaatgaccgaaaattagtagaatacttaaaggagagt 420

Qy    439 tcttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaatat 498
          |||||||||||| |||||||| ||||||||||||  |||||  ||||||||||||||||||
Db    421 tcttttgatgcggtgtttcttgatccttttgatgcctgtgcgttaattgttgccaaatat 480

Qy    499 ttctccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaaggt 558
          |||||||||||||| ||||||||||||||||||||  ||||||||||||||||||||||
Db    481 ttctccctcccctctgtggtcttcgccaggggaataggttgccactatcttgaagaaggt 540

Qy    559 gcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgcc 618
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 gcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgcc 600

Qy    619 atgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatgc 678
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 atgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattttgc 660

Qy    679 caccgtttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttacg 738
          ||   ||||| ||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    661 cagtatttttccaaaaatgccctagaaatagcctctgaaattctccaaacacctgtcaca 720

Qy    739 gagtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttggac 798
          |  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 gcatatgatctctacagccacacatcaatttggttgttgcgaacagactttgttttggac 780

Qy    799 tatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggga 858
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    781 tatcccaaacccgtgatgcccaatatgatcttcattggtggtatcaactgccatcaggga 840

Qy    859 aagccgttgcctatggaatttgaagcctacattaatgcttctgga 903
          ||||| |||||||||| |  | |   || | |||   | |  |||
Db    841 aagccattgcctatggtaagtcacctctcctttagcacattagga 885


RESULT  12
AAZ45115
ID   AAZ45115 standard; DNA; 951 BP.
XX
AC   AAZ45115;
XX
DT   28-FEB-2000  (first entry)
XX
```

```
Qy   2010 aattgatccttaaccaataatggtcagtcctcatctctgtcctgcttcataggtgccacc 2069
          ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    980 aattgatccttaaccaataatggtcagtcctcatctctgtcgtgcttcataggtgccacc 1039

Qy   2070 ttgtgtgtttaaagaagggaagctttgacctttagagtgtag---------------- 2112
          ||||||||||||||||||||||||||||||||||||||||
Db   1040 ttgtgtgtttaaagaagggaagctttgacctttagagtgtagtgaaatgaatgaatggc 1099

Qy   2113 ------------------------------------------------------------ 2112

Db   1100 ttggagtgcactgagaacagcatatgatttcttgctttggggaaaaagaatgatgctatg 1159

Qy   2113 ------------------gtgaaatgaatgaatggcttggagtgcactgagaacagcata 2154
                            ||||||||||||||||||||||||||||||||||||||||||
Db   1160 aaattggtgggtggtgtagtgaaatgaatgaatggcttggagtgcactgagaacagcata 1219

Qy   2155 tgatttcttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgag 2214
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1220 tgatttcttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgag 1279

Qy   2215 aagataatcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagcta 2274
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1280 aagataatcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagcta 1339

Qy   2275 tgaagtgctgggcaagtttacttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||||||||||
Db   1340 tgaagtgctgggcaagtttacttttttctgatgtttcctacaact 1385
```

RESULT 10
AAH51405
ID AAH51405 standard; DNA; 1001 BP.
XX
AC AAH51405;
XX
DT 29-AUG-2001 (first entry)
XX
DE Human UGT1A7 related DNA containing a biallelic polymorphism SEQ ID 296.
XX
KW Human; biallelic marker; single nucleotide polymorphism; SNP; MGSTII;
KW microsomal glutathione S-transferase II; malate decarboxylase enzyme;
KW DME1; ME1; cytochrome P450; glutathione reductase; GSHR; GSHS; GGT5;
KW flavin-containing monooxygenase; FMO; gamma-glutamyltransferase 5;
KW dipeptidase; DP; glucose 6-phosphate dehydrogenase; G6PDH; haplotype;
KW phosphogluconate dehydrogenase; PGDH; drug metabolism; phenotype;
KW uridine diphosphate glucoronosyl transferase; UGT2; asthma; hepatoxicity;
KW zileuton; ds.
XX
OS Homo sapiens.
XX
PN WO200058508-A2.
XX
PD 05-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-IB00403.
XX
PR 25-MAR-1999; 99US-0126269.
PR 30-APR-1999; 99US-0131961.
XX
PA (GEST) GENSET.
XX
PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A;
XX
DR WPI; 2000-638353/61.
XX
PT Polynucleotides comprising sequences from malate decarboxylase
PT enzyme-related biallelic markers used for genotyping -
XX
PS Claim 13; Page 488-489; 673pp; English.
XX
CC Sequences AAH51110-AAH51593 represent human DNA fragments which contain
CC biallelic markers. The sequences are related to various human genes
CC including microsomal glutathione S-transferase II (MGSTII), malate
CC decarboxylase enzyme (DME1/ME1), cytochrome P450, glutathione
CC reductase/synthase (GSHR/GSHS), flavin-containing monooxygenases (FMO),
CC gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 6-phosphate
CC dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine
CC diphosphate glucoronosyl transferases (UGT2). Each of these sequences
CC contains a biallelic marker/polymorphism, which is represented in the
CC sequence as a degenerate/undefined base. The genes to which the biallelic
CC marker containing sequences are related are involved in drug metabolism.
CC Sequences AAH51594 - AAH51598 represent the genomic sequence of the
CC MGSTII gene and four alternative MGSTII cDNA sequences. AAB62905-AAB62906
CC are MGSTII gene products. PCR primers AAH51599 and AAH51600 are used in
CC an example for the amplification of human genomic DNA fragments. The
CC invention includes a method of genotyping comprising determining the
CC identity of a nucleotide at a DME- or MGSTII-related biallelic marker in
CC a biological sample. The method is used to determine the frequency in
CC population of an allele of a DME- or MGST-II related biallelic marker and
CC to select an individual for inclusion in a clinical trial of a drug
CC treatment. The method is also used to detect association between allele
CC and phenotype, and to detect association between haplotype and phenotype.
CC The polynucleotides are used, in hybridization assays, sequencing assays
CC or allele specific amplification assays. The method can be used to
CC determine whether an individual suffers or is at risk of developing
CC asthma or is at risk of developing hepatoxicity on treatment with
CC zileuton.
XX
SQ Sequence 1001 BP; 315 A; 169 C; 205 G; 311 T; 1 other;

Query Match 33.2%; Score 769.6; DB 21; Length 1001;
Best Local Similarity 99.9%; Pred. No. 3e-208;
Matches 769; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1551 ccggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattg 1610
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ccggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattg 60

Qy   1611 agaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttccaaacttga 1670
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 agaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttccaaacttga 120

Qy   1671 aaacagaatcagtgttaaattcattttattcttattaaggaaatactttgcataaattaa 1730
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 aaacagaatcagtgttaaattcattttattcttattaaggaaatactttgcataaattaa 180

Qy   1731 tcagccccagagtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagt 1790
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 tcagccccagagtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagt 240

Qy   1791 aaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgcc 1850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 aaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgcc 300

Qy   1851 atttttcagaggacgtgcagacaggctggcattctagattacttttcttactctgaaaca 1910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 atttttcagaggacgtgcagacaggctggcattctagattacttttcttactctgaaaca 360

Qy   1911 tggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatgg 1970
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 tggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatgg 420

Qy   1971 cagttttaatcttatcttttggcttctgcagatggttgcaattgatccttaaccaataat 2030
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 cagttttaatcttatcttttggcttctgcagatggttgcaattgatccttaaccaataat 480

Qy   2031 ggtcagtcctcatctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaa 2090
          ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db    481 ggtcagtcctcatctctgtcstgcttcataggtgccaccttgtgtgtttaaagaagggaa 540

Qy   2091 gctttgtacctttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacag 2150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 gctttgtacctttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacag 600

Qy   2151 catatgatttcttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatt 2210
```

```
Db    541 tggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatctctgt 600

Qy   2050 cctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacctttagagtg 2109
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 cctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacctttagagtg 660

Qy   2110 taggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttcttgctttg 2169
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 taggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttcttgctttg 720

Qy   2170 gggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatcattgct 2229
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 gggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatcattgct 780

Qy   2230 tatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgctgggcaa 2289
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 tatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgctgggcaa 840

Qy   2290 gtttactttttttctgatgtttcctacaact 2320
          |||| |||||||||||||||||||||||||
Db    841 gtttactttttttctgatgtttcctacaact 871


RESULT   9
AAQ33027
ID   AAQ33027 standard; DNA; 1667 BP.
XX
AC   AAQ33027;
XX
DT   27-JAN-1993  (first entry)
XX
DE   Exon 5 from the UGT1 gene locus.
XX
KW   UGT1A; UGT1BP; UGT1C; UGT1D; UGT1E; UGT1F; isozyme; bilirubin;
KW   UDP-glucuronosyl transferase; CN; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        1..20
FT                   /*tag=  a
FT                   /note= "representation of 3.3 kbp of
FT                          non-sequenced DNA between the sequences
FT                          represented in AAQ33026 and AAQ33027"
FT   exon            291..1402
FT                   /*tag=  b
FT                   /label= Exon_5
FT   misc_RNA        291..586
FT                   /*tag=  c
FT                   /note= "encodes part of transferase C-terminal; see CC"
FT   polyA_signal    1389..1394
FT                   /*tag=  d
XX
PN   WO9212987-A.
XX
PD   06-AUG-1992.
XX
PF   10-JAN-1992;   92WO-US00282.
XX
PR   10-JAN-1991;   91US-0639453.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Owens IS,  Ritter JK;
XX
DR   WPI; 1992-284593/34.
DR   P-PSDB; AAR30166.
XX
PT   Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT   for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT   types I and II
XX
PS   Disclosure; Fig 1I; 99pp; English.

XX
CC   The isolated gene locus, UGT1, has a sequence of about 10000 bp
CC   which represent (1) Exon 1, comprising 6 transcriptional units
CC   (UGT1F, E, D, C, BP and A), represented in AAQ27368 and
CC   AAQ33020-24 respectively;
CC                  (2) Exon 2, represented in AAQ33025;
CC                  (3) Exon 3, represented in AAQ33026;
CC                  (4) Exon 4, represented in AAQ33026;
CC                  (5) Exon 5, represented in AAQ33027; and
CC                  (6) about 69 kb of non-sequenced DNA.
CC   Six unique N-termini of 286-289 amino acids are encoded by
CC   the six different first exons and identical C-termini of 246 amino
CC   acids are encoded by the common exons 2-5.  The UGT1 gene locus
CC   encodes a family of UDP-glucuronosyl transferase isozymes, two of
CC   which metabolise bilirubin.
CC   Patients having Crigler-Najjar Syndrome (CN) Type I, have a
CC   mutation present in the second common exon.
XX
SQ   Sequence 1667 BP; 487 A; 316 C; 370 G; 474 T; 20 other;


  Query Match             34.4%;  Score 798;  DB 13;  Length 1667;
  Best Local Similarity   89.1%;  Pred. No. 3.4e-216;
  Matches 1003;  Conservative    0;  Mismatches   20;  Indels  103;  Gaps    9;

Qy   1291 aaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctccagccttcac 1350
          || | |    ||  ||     ||||||||||||||||||||||||||||||||||||||||
Db    267 aacgaaactgtctttgtgtttagttacaaggagaacatcatgcgcctctccagccttcac 326

Qy   1351 aaggaccgcccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgagg 1410
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    327 aaggaccgcccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgagg 386

Qy   1411 cacaagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccat 1470
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    387 cacaagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccat 446

Qy   1471 tccttggacgtgattggtttcctcttggccgtcgtgct-gacagtggccttcatcacctt 1529
          ||||||||||||||||||||||||||||||||||  ||| ||||||||||||||||||||
Db    447 tccttggacgtgattggtttcctcttggccgtctggctggacagtggccttcatcacctt 506

Qy   1530 taaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcgagttaagaaagc 1589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    507 taaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcgagttaagaaagc 566

Qy   1590 ccacaaatccaagacccattgagaagtgggtgggaaataaggtaaaattttgaaccattc 1649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    567 ccacaaatccaagacccattgagaagtgggtgggaaataaggtaaaattttgaaccattc 626

Qy   1650 cctagtcatttccaaacttgaaaacagaatcagtgttaaattcattttattcttattaag 1709
          |||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db    627 cctagtcatttcc-aacttgaaaacagaatcagtgttaaattcattttattcttattaag 685

Qy   1710 gaaatactttgcataaattaatcagccccagagtgctttaaaaaattctctttaaataaaa 1769
          |||||||  |||||||||||||||||||||  ||||||||||||||||||||||||||||
Db    686 gaaatac-ttgcataaattaatcagcccc-gagtgctttaaaaaattctctttaaataaaa 743

Qy   1770 ataatagactcgctagtcagtaaagatatttgaatatgtatcgtgccccctccggtgtct 1829
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||| |||
Db    744 ataatagactcgctagtcagtaaagatatttgaatatgtatcgtgccccctctggtctct 803

Qy   1830 ttgatcaggatgacatgtgccatttttcagaggacgtgcagacaggctggcattctagat 1889
          |||||||| ||||||||||||||||||||||||||||||||||||||| |||||||||
Db    804 ttgatcagcatgacatgtgccatttttcagaggacgtgcagacaggctgg-attctagat 862

Qy   1890 tacttttcttactctgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccac 1949
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    863 tacttttcttactctgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccac 922

Qy   1950 cgctgcccctactgcaaatggcagttttaatcttatcttttggcttctgcagatggttgc 2009
          |||  ||||||||||||||||||  ||||||||||||||||||||||| |||||||
Db    923 ggct-cccctactgcaaatggcag-tttaatcttatcttttggcttctgcag-tggttgc 979
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 gacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttg 600

Qy   1922 ggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatc 1981
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatc 660

Qy   1982 ttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctc 2041
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctc 720

Qy   2042 atctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacct 2101
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 atctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacct 780

Qy   2102 ttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttc 2161
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttc 840

Qy   2162 ttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataa 2221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataa 900

Qy   2222 tcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtg 2281
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 tcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtg 960

Qy   2282 ctgggcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||||||
Db    961 ctgggcaagtttacttttttttctgatgtttcctacaact 999
```

RESULT 8
AAH51404
ID AAH51404 standard; DNA; 1001 BP.
XX
AC AAH51404;
XX
DT 29-AUG-2001 (first entry)
XX
DE Human UGT1A7 related DNA containing a biallelic polymorphism SEQ ID 295.
XX
KW Human; biallelic marker; single nucleotide polymorphism; SNP; MGSTII;
KW microsomal glutathione S-transferase II; malate decarboxylase enzyme;
KW DME1; ME1; cytochrome P450; glutathione reductase; GSHR; GSHS; GGT5;
KW flavin-containing monooxygenase; FMO; gamma-glutamyltransferase 5;
KW dipeptidase; DP; glucose 6-phosphate dehydrogenase; G6PDH; haplotype;
KW phosphogluconate dehydrogenase; PGDH; drug metabolism; phenotype;
KW uridine diphosphate glucoronosyl transferase; UGT2; asthma; hepatoxicity;
KW zileuton; ds.
XX
OS Homo sapiens.
XX
PN WO200058508-A2.
XX
PD 05-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-IB00403.
XX
PR 25-MAR-1999; 99US-0126269.
PR 30-APR-1999; 99US-0131961.
XX
PA (GEST) GENSET.
XX
PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A;
XX
DR WPI; 2000-638353/61.
XX
PT Polynucleotides comprising sequences from malate decarboxylase
PT enzyme-related biallelic markers used for genotyping -
XX
PS Claim 13; Page 487-488; 673pp; English.
XX
CC Sequences AAH51110-AAH51593 represent human DNA fragments which contain
CC biallelic markers. The sequences are related to various human genes
CC including microsomal glutathione S-transferase II (MGSTII), malate
CC decarboxylase enzyme (DME1/ME1), cytochrome P450, glutathione
CC reductase/synthase (GSHR/GSHS), flavin-containing monooxygenases (FMO),
CC gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 6-phosphate
CC dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine
CC diphosphate glucoronosyl transferases (UGT2). Each of these sequences
CC contains a biallelic marker/polymorphism, which is represented in the
CC sequence as a degenerate/undefined base. The genes to which the biallelic
CC marker containing sequences are related are involved in drug metabolism.
CC Sequences AAH51594 - AAH51598 represent the genomic sequence of the
CC MGSTII gene and four alternative MGSTII cDNA sequences. AAB62905-AAB62906
CC are MGSTII gene products. PCR primers AAH51599 and AAH51600 are used in
CC an example for the amplification of human genomic DNA fragments. The
CC invention includes a method of genotyping comprising determining the
CC identity of a nucleotide at a DME- or MGSTII-related biallelic marker in
CC a biological sample. The method is used to determine the frequency in
CC population of an allele of a DME- or MGST-II related biallelic marker and
CC to select an individual for inclusion in a clinical trial of a drug
CC treatment. The method is also used to detect association between allele
CC and phenotype, and to detect association between haplotype and phenotype.
CC The polynucleotides are used, in hybridization assays, sequencing assays
CC or allele specific amplification assays. The method can be used to
CC determine whether an individual suffers or is at risk of developing
CC asthma or is at risk of developing hepatoxicity on treatment with
CC zileuton.
XX
SQ Sequence 1001 BP; 293 A; 171 C; 210 G; 326 T; 1 other;

Query Match 37.5%; Score 870.6; DB 21; Length 1001;
Best Local Similarity 99.9%; Pred. No. 5.6e-237;
Matches 870; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1450 ctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcgtgctg 1509
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcgtgctg 60

Qy   1510 acagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaa 1569
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 acagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaa 120

Qy   1570 aaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtgggaaataa 1629
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 aaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtgggaaataa 180

Qy   1630 ggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgttaaa 1689
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgttaaa 240

Qy   1690 ttcattttattcttattaaggaaatactttgcataaattaatcagccccagagtgcttta 1749
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ttcattttattcttattaaggaaatactttgcataaattaatcagccccagagtgcttta 300

Qy   1750 aaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaatatgta 1809
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 aaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaatatgta 360

Qy   1810 tcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggacgtgca 1869
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 tcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggacgtgca 420

Qy   1870 gacaggctggcattctagattacttttcttactctgaaacatggcctgtttgggagtgcg 1929
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 gacaggctggcattctagattacttttcttactctgaaacatggcctgtttgggagtgcg 480

Qy   1930 ggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatcttatcttt 1989
          |||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db    481 ggattcaaaggtggtcccacsgctgcccctactgcaaatggcagttttaatcttatcttt 540

Qy   1990 tggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatctctgt 2049
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy    1891 acttttcttactctgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccacc 1950
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15167 acttttcttactctgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccacg 15226

Qy    1951 gctgcccctactgcaaatggcagttttaatcttatcttttggcttctgcagatggttgca 2010
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15227 gctgcccctactgcaaatggcagttttaatcttatcttttggcttctgcagatggttgca 15286

Qy    2011 attgatccttaaccaataatggtcagtcctcatctctgtcctgcttcataggtgccacct 2070
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15287 attgatccttaaccaataatggtcagtcctcatctctgtcgtgcttcataggtgccacct 15346

Qy    2071 tgtgtgtttaaagaagggaagctttgtacctttagagtgtaggtgaaatgaatgaatggc 2130
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15347 tgtgtgtttaaagaagggaagctttgtacctttagagtgtaggtgaaatgaatgaatggc 15406

Qy    2131 ttggagtgcactgagaacagcatatgatttcttgctttggggaaaaagaatgatgctatg 2190
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15407 ttggagtgcactgagaacagcatatgatttcttgctttggggaaaaagaatgatgctatg 15466

Qy    2191 aaattggtgggtggtgtatttgagaagataatcattgcttatgtcaaatggagctgaatt 2250
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15467 aaattggtgggtggtgtatttgagaagataatcattgcttatgtcaaatggagctgaatt 15526

Qy    2251 tgataaaaacccaaaatacagctatgaagtgctgggcaagtttacttttttttctgatgtt 2310
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15527 tgataaaaacccaaaatacagctatgaagtgctgggcaagtttacttttttttctgatgtt 15586

Qy    2311 tcctacaact 2320
           ||||||||||
Db   15587 tcctacaact 15596
```

```
RESULT   7
AAH51403
ID   AAH51403 standard; DNA; 1001 BP.
XX
AC   AAH51403;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Human UGT1A7 related DNA containing a biallelic polymorphism SEQ ID 294.
XX
KW   Human; biallelic marker; single nucleotide polymorphism; SNP; MGSTII;
KW   microsomal glutathione S-transferase II; malate decarboxylase enzyme;
KW   DME1; ME1; cytochrome P450; glutathione reductase; GSHR; GSHS; GGT5;
KW   flavin-containing monooxygenase; FMO; gamma-glutamyltransferase 5;
KW   dipeptidase; DP; glucose 6-phosphate dehydrogenase; G6PDH; haplotype;
KW   phosphogluconate dehydrogenase; PGDH; drug metabolism; phenotype;
KW   uridine diphosphate glucoronosyl transferase; UGT2; asthma; hepatoxicity;
KW   zileuton; ds.
XX
OS   Homo sapiens.
XX
PN   WO200058508-A2.
XX
PD   05-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-IB00403.
XX
PR   25-MAR-1999;   99US-0126269.
PR   30-APR-1999;   99US-0131961.
XX
PA   (GEST ) GENSET.
XX
PI   Blumenfeld M,  Bougueleret L,  Chumakov I,  Cohen-Akenine A;
XX
DR   WPI; 2000-638353/61.
XX
PT   Polynucleotides comprising sequences from malate decarboxylase
PT   enzyme-related biallelic markers used for genotyping -
XX
PS   Claim 13; Page 487; 673pp; English.
```

```
XX
CC   Sequences AAH51110-AAH51593 represent human DNA fragments which contain
CC   biallelic markers. The sequences are related to various human genes
CC   including microsomal glutathione S-transferase II (MGSTII), malate
CC   decarboxylase enzyme (DME1/ME1), cytochrome P450, glutathione
CC   reductase/synthase (GSHR/GSHS), flavin-containing monooxygenases (FMO),
CC   gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 6-phosphate
CC   dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine
CC   diphosphate glucoronosyl transferases (UGT2). Each of these sequences
CC   contains a biallelic marker/polymorphism, which is represented in the
CC   sequence as a degenerate/undefined base. The genes to which the biallelic
CC   marker containing sequences are related are involved in drug metabolism.
CC   Sequences AAH51594 - AAH51598 represent the genomic sequence of the
CC   MGSTII gene and four alternative MGSTII cDNA sequences. AAB62905-AAB62906
CC   are MGSTII gene products. PCR primers AAH51599 and AAH51600 are used in
CC   an example for the amplification of human genomic DNA fragments. The
CC   invention includes a method of genotyping comprising determining the
CC   identity of a nucleotide at a DME- or MGSTII-related biallelic marker in
CC   a biological sample. The method is used to determine the frequency in
CC   population of an allele of a DME- or MGST-II related biallelic marker and
CC   to select an individual for inclusion in a clinical trial of a drug
CC   treatment. The method is also used to detect association between allele
CC   and phenotype, and to detect association between haplotype and phenotype.
CC   The polynucleotides are used, in hybridization assays, sequencing assays
CC   or allele specific amplification assays. The method can be used to
CC   determine whether an individual suffers or is at risk of developing
CC   asthma or is at risk of developing hepatoxicity on treatment with
CC   zileuton.
XX
SQ   Sequence 1001 BP; 269 A; 205 C; 233 G; 293 T; 1 other;
```

```
  Query Match             43.0%;  Score 998.6;  DB 21;  Length 1001;
  Best Local Similarity   99.9%;  Pred. No. 2.1e-273;
  Matches  998;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1322 agaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctgg 1381
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 agaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctgg 60

Qy    1382 ccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcag 1441
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcag 120

Qy    1442 cccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccg 1501
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 cccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccg 180

Qy    1502 tcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgct 1561
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 tcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgct 240

Qy    1562 tggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtg 1621
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 tggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtg 300

Qy    1622 ggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatca 1681
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatca 360

Qy    1682 gtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccaga 1741
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 gtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccaga 420

Qy    1742 gtgctttaaaaaattctcttaaataaaaaataatagactcgctagtcagtaaagatatttg 1801
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 gtgctttaaaaaattctcttaaataaaaaataatagactcgctagtcagtaaagatatttg 480

Qy    1802 aatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagag 1861
           ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db     481 aatatgtatcgtgccccctcyggtgtctttgatcaggatgacatgtgccatttttcagag 540

Qy    1862 gacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttg 1921
```

```
XX
FH   Key             Location/Qualifiers
FT   variation       replace (2510,C)
FT                   /*tag= a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   CDS             2616..14887
FT                   /*tag= b
FT                   /product= "UDP glycosyltransferase 1"
FT                   /note= "This CDS is specifically claimed in claim
FT                         26"
FT   variation       replace (2756,T)
FT                   /*tag= c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (2826,A)
FT                   /*tag= d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (3155,G)
FT                   /*tag= e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (9508,C)
FT                   /*tag= f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (9511,T)
FT                   /*tag= g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (10091,T)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (10094,T)
FT                   /*tag= i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (10095,C)
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (10140,C)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (14423,T)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (14713,T)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (14776,T)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (14971,C)
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   WO200179230-A2.
XX
PD   25-OCT-2001.
XX
PF   13-APR-2001; 2001WO-US12273.
XX
PR   18-APR-2000; 2000US-197514P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Chew A,  Choi JY,  Koshy B,  Rounds E;
XX
DR   WPI; 2002-075063/10.
DR   P-PSDB; AAU10940.
XX
PT   Genotyping a human UDP glycosyltransferase 1 gene of an individual for
PT   determining the haplotype of an individual, involves determining the
PT   identity of a nucleotide pair at specific polymorphic sites for two
PT   copies of the gene -
XX
PS   Claim 21; Fig 1; 81pp; English.
XX
CC   The invention relates to genotyping a human UDP glycosyltransferase
CC   (UGT1A1) gene of an individual, involving determining for the two copies
CC   of the UGT1A1 gene present in the individual, the identity of the
CC   nucleotide pair at one or more polymorphic sites. The new method is
CC   useful for determining whether an individual has a haplotype or haplotype
CC   pairs, given in the specification. It is useful for improving the
CC   efficacy and reliability of several steps in the discovery and
CC   development of drugs for treating diseases associated with UGT1A1
CC   activity, e.g., Gilbert's syndrome and Crigler-Najjar syndrome, to
CC   validate UGT1A1 as a candidate agent for treating a specific condition or
CC   disease predicted to be associated with UGT1A1 activity, and in the
CC   design of clinical trials of candidate drugs for treating a specific
CC   condition or disease predicted to be associated with UGT1A1 activity. The
CC   method is useful to screen for compounds targeting UGT1A1 to treat a
CC   specific condition or disease associated with UGT1A1 activity. A nucleic
CC   acid (I) comprising a polymorphic variant of a reference sequence for
CC   the UGT1A1 gene or cDNA (II) or its fragment is useful in studying the
CC   expression and function of UGT1A1, and in expressing UGT1A1 protein for
CC   use in screening for candidate drugs to treat diseases related to UGT1A1
CC   activity. (I) or (II) is useful for therapeutic purposes. (II) or a
CC   recombinant organism comprising (II) is useful for studying
CC   expression of the UGT1A1 isogenes in vivo, for in vivo screening and
CC   testing of drugs targeted against UGT1A1 protein, and for testing the
CC   efficacy of therapeutic agents and compounds for Gilbert's syndrome and
CC   Crigler-Najjar syndrome, in a biological system. The present sequence
CC   represents the coding sequence of UDP glycosyltransferase 1.
XX
SQ   Sequence 18887 BP; 5079 A; 4037 C; 4131 G; 5640 T; 0 other;


  Query Match             43.3%;  Score 1004.4;  DB 24;  Length 18887;
  Best Local Similarity   98.4%;  Pred. No. 2.8e-274;
  Matches 1014;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy    1291 aaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctccagccttcac 1350
           || | |    ||  ||     ||||||||||||||||||||||||||||||||||||||||||
Db   14567 aacgaaactgtctttgtgtttagttacaaggagaacatcatgcgcctctccagccttcac 14626

Qy    1351 aaggaccgcccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgagg 1410
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14627 aaggaccgcccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgagg 14686

Qy    1411 cacaagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccat 1470
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14687 cacaagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccat 14746

Qy    1471 tccttggacgtgattggtttcctcttggccgtcgtgctgacagtggccttcatcaccttt 1530
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14747 tccttggacgtgattggtttcctcttggccgtcgtgctgacagtggccttcatcaccttt 14806

Qy    1531 aaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcgagttaagaaagcc 1590
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14807 aaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcgagttaagaaagcc 14866

Qy    1591 cacaaatccaagacccattgagaagtgggtgggaaataaggtaaaattttgaaccattcc 1650
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14867 cacaaatccaagacccattgagaagtgggtgggaaataaggtaaaattttgaaccattcc 14926

Qy    1651 ctagtcatttccaaacttgaaaacagaatcagtgttaaattcattttattcttattaagg 1710
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14927 ctagtcatttccaaacttgaaaacagaatcagtgttaaattcattttattcttattaagg 14986

Qy    1711 aaatactttgcataaattaatcagccccagagtgctttaaaaaattctcttaaataaaaa 1770
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14987 aaatactttgcataaattaatcagccccagagtgctttaaaaaattctcttaaataaaaa 15046

Qy    1771 taatagactcgctagtcagtaaagatatttgaatatgtatcgtgccccctccggtgtctt 1830
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   15047 taatagactcgctagtcagtaaagatatttgaatatgtatcgtgccccctctggtgtctt 15106

Qy    1831 tgatcaggatgacatgtgccatttttcagaggacgtgcagacaggctggcattctagatt 1890
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15107 tgatcaggatgacatgtgccatttttcagaggacgtgcagacaggctggcattctagatt 15166
```

CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1655 BP; 402 A; 388 C; 403 G; 462 T; 0 other;

Query Match 66.9%; Score 1551.8; DB 23; Length 1655;
Best Local Similarity 97.7%; Pred. No. 0;
Matches 1574; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
Qy      1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     45 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 104

Qy     61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    105 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 164

Qy    121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcagggggcatgag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    165 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcagggggcatgag 224

Qy    181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    225 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 284

Qy    241 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    285 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 344

Qy    301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360
          |||||||||||||||||||||||||||||| | | |||||||||  || ||||||  ||||
Db    345 catgctcaatggaaagcacaagtacgaagtttgttttctctatttctgagttcatccaat 404

Qy    361 gacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagta 420
          |    |||||  ||||||||||||||  ||||||||||||||||||  |||  ||||||||||
Db    405 ggtttttttaacttatttttttcgcattgcaggagtttgtttaatgaccgaaaattagta 464

Qy    421 gaatacttaaaggagagttctttgatgcagtgtttctcgatccttttgataactgtggc 480
          |||||||||||||||||||||||||||||| ||||||| ||||||||||||  |||||
Db    465 gaatacttaaaggagagttctttgatgcggtgtttcttgatccttttgatgcctgtgcg 524

Qy    481 ttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgc 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db    525 ttaattgttgccaaatatttctccctcccctctgtggtcttcgccaggggaataggttgc 584

Qy    541 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    585 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 644

Qy    601 ttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    645 ttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttg 704

Qy    661 gaggaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaatt 720
          |||||||||||||| |||||   ||||| ||||||||||||||||||||||||||||||
Db    705 gaggaacatttatttgccagtatttttccaaaaatgccctagaaatagcctctgaaatt 764

Qy    721 ctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcga 780
          |||||||||||||| || |  ||||||||||||||||||||||||||||||||||||||
Db    765 ctccaaacacctgtcacagcatatgatctctacagccacacatcaatttggttgttgcga 824

Qy    781 acggacttttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggt 840
          || |||||  |||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db    825 acagacttttgttttggactatcccaaacccgtgatgcccaatatgatcttcattggtggt 884

Qy    841 atcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttct 900
          ||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db    885 atcaactgccatcagggaaagccattgcctatggaatttgaagcctacattaatgcttct 944

Qy    901 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    945 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 1004

Qy    961 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1005 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1064

Qy   1021 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1065 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaa 1124

Qy   1081 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1125 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1184

Qy   1141 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1185 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1244

Qy   1201 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1245 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1304

Qy   1261 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1305 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1364

Qy   1321 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1365 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1424

Qy   1381 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1425 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1484

Qy   1441 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1485 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1544

Qy   1501 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1545 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1604

Qy   1561 ttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattga 1611
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1605 ttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattga 1655
```

RESULT 6
AAS18543
ID AAS18543 standard; DNA; 18887 BP.
XX
AC AAS18543;
XX
DT 12-MAR-2002 (first entry)
XX
DE DNA encoding UDP glycosyltransferase 1 (UGT1A1).
XX
KW UDP glycosyltransferase 1; UGT1A1; human; haplotyping;
KW drug discovery; Gilbert's syndrome; Crigler-Najjar syndrome; ds.
XX
OS Homo sapiens.

```
Qy    858 aaagccgttgcctatggaatttgaagcctacattaatgcttctggagaacatggaattgt 917
Db    881 ggaagccactatctaggaatttgaagcctacattaatgcttctggagaacatggaattgt 940

Qy    918 ggttttctctttgggatcaatggtctcagaaattccagagaagaaagctatggcaattgc 977
Db    941 ggttttctctttggaatcaatggtctcagaaattccagagaagaaagctatggcaattgc 1000

Qy    978 tgatgctttgggcaaaatccctcagacagtcctgtggcggtacactggaacccgaccatc 1037
Db   1001 tgatgctttgggcaaaatccctcagacagtcctgtggcggtacactggaacccgaccatc 1060

Qy   1038 gaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacgatctgcttggtca 1097
Db   1061 gaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacgatctgcttggtca 1120

Qy   1098 cccgatgacccgtgcctttatcacccatgctggttcccatggtgtttatgaaagcatatg 1157
Db   1121 cccgatgacccgtgcctttatcacccatgctggttcccatggtgtttatgaaagcatatg 1180

Qy   1158 caatggcgttcccatggtgatgatgcccttgtttggtgatcagatggacaatgcaaagcg 1217
Db   1181 caatggcgttcccatggtgatgatgcccttgtttggtgatcagatggacaatgcaaagcg 1240

Qy   1218 catggagactaagggagctggagtgaccctgaatgttctggaaatgacttctgaagattt 1277
Db   1241 catggagactaagggagctggagtgaccctgaatgttctggaaatgacttctgaagattt 1300

Qy   1278 agaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcct 1337
Db   1301 agaaaatgctcaaaaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcct 1360

Qy   1338 ctccagccttcacaaggaccgcccggtggagccgctggacctggccgtgttctgggtgga 1397
Db   1361 ctccagccttcacaaggaccgcccggtggagccgctggacctggccgtgttctgggtgga 1420

Qy   1398 gtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagcccacgacctcacctg 1457
Db   1421 gtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagcccacgacctcacctg 1480

Qy   1458 gtaccagtaccattccttggacgtgattggtttcctcttggccgtcgtgctgacagtggc 1517
Db   1481 gtaccagtaccattccttggacgtgattggtttcctcttggccgtgctgctgacagtggc 1540

Qy   1518 cttcatcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcg 1577
Db   1541 cttcatcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcg 1600

Qy   1578 agttaagaaagcccacaaatccaagacccattgagaagtgggtgggaaataaggtaaaat 1637
Db   1601 agttaagaaagcccacaaatccaagacccattgagaagtgggtgggaaataaggtaaaat 1660

Qy   1638 tttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgttaaattcatttt 1697
Db   1661 tttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgttaaattcatttt 1720

Qy   1698 attcttattaaggaaatactttgcataaattaatcagccccagagtgctttaaaaaattc 1757
Db   1721 attcttattaaggaaatactttgcat-aattaatcagccccagagtgctttaaaaaattc 1779

Qy   1758 tcttaaataaaaataatagactcgctagtcagtaaagatatttgaatatgtatcgtgccc 1817
Db   1780 tcttaaataaaaataatagactcgctagtcagtaaagatatttgaatatgtatcgtgccc 1839

Qy   1818 cctccggtgtctttgatcaggatgacatgtgccatttttcagaggacgtgcagacaggct 1877
Db   1840 cctccggtgtctttgatcaggatgacatgtgccatttttcagaggacgtgcagacaggct 1899

Qy   1878 ggcattctagattacttttcttactctgaaacatggcctgtttgggagtgcgggattcaa 1937
Db   1900 ggcattctagattacttttcttactctgaaacatggcctgtttgggagtgcgggattcaa 1959

Qy   1938 aggtggtcccaccgctgcccctactgcaaatggcagttttaatcttatcttttggcttct 1997
Db   1960 aggtggtcccaccgctgcccctactgcaaatggcagttttaatcttatcttttggcttct 2019

Qy   1998 gcagatggttgcaattgatccttaaccaataatggtcagtcctcatctctgtcctgcttc 2057
Db   2020 gcagatggttgcaattgatccttaaccaataatggtcagtcctcatctctgtcctgcttc 2079

Qy   2058 ataggtgccacc-ttgtgtgtttaaagaagggaagctttgtacctttagagtgtaggtga 2116
Db   2080 ataggtgccacctttgtgtgtttaaagaagggaagctttgtacctttagagtgtaggtga 2139

Qy   2117 aatgaatgaatggcttggagtgcactgagaacagcatatgatttcttgctttggggaaaa 2176
Db   2140 aatgaatgaatggcttggagtgcactgagaacagcatatgatttcttgctttggggaaaa 2199

Qy   2177 agaatgatgctatgaaattggtgggtggtgtatttgagaagataatcattgcttatgtca 2236
Db   2200 agaatgatgctatgaaattggtgggtggtgtatttgagaagataatcattgcttatgtca 2259

Qy   2237 aatggagctgaatttgataaaaacccaaaatacagctatgaagtgctgggcaagtttact 2296
Db   2260 aatggagctgaatttgataaaaacccaaaatacagctatgaagtgctgggcaagtttact 2319

Qy   2297 ttttttctgatgtttcctacaact 2320
Db   2320 ttttttctgatgtttcctacaact 2343
```

RESULT 5
AAS91252
ID AAS91252 standard; cDNA; 1655 BP.
XX
AC AAS91252;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #27056.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG27065.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 27056; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques

```
Db   1870 gtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttggga 1929

Qy   1925 gtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatctta 1984
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1930 gtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatctta 1989

Qy   1985 tctttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatc 2044
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1990 tctttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatc 2049

Qy   2045 tctgtcctgcttcataggtgccacc-ttgtgtgtttaaagaagggaagctttgtacctttt 2103
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   2050 tctgtcctgcttcataggtgccacctttgtgtgtttaaagaagggaagctttgtacctttt 2109

Qy   2104 agagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttctt 2163
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2110 agagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttctt 2169

Qy   2164 gctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatc 2223
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2170 gctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatc 2229

Qy   2224 attgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgct 2283
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2230 attgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgct 2289

Qy   2284 gggcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||||
Db   2290 gggcaagtttacttttttttctgatgtttcctacaact 2326


RESULT   4
AAQ27370
ID   AAQ27370 standard; cDNA; 2368 BP.
XX
AC   AAQ27370;
XX
DT   27-JAN-1993  (first entry)
XX
DE   HUG-Br2.
XX
KW   Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
KW   monoglucuronide; diglucuronide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             30..801
FT                   /*tag= a
FT   polyA_signal    2347..2352
FT                   /*tag= b
FT                   /number= 1
FT   polyA_signal    2355..2360
FT                   /*tag= c
FT                   /number= 2
XX
PN   WO9212987-A.
XX
PD   06-AUG-1992.
XX
PF   10-JAN-1992;   92WO-US00282.
XX
PR   10-JAN-1991;   91US-0639453.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Owens IS,  Ritter JK;
XX
DR   WPI; 1992-284593/34.
DR   P-PSDB; AAR26154.
XX
PT   Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT   for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT   types I and II
XX
PS   Disclosure; Fig 9A-I; 99pp; English.
XX
CC   Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
CC   been isolated.  They are referred to as HUGBr1 (AAQ27369) and HUGBr2
CC   (AAQ27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
CC   upon expression individually in COS-1 cells, encode isoforms that
CC   catalyse the formation of the two bilirubin monoglucuronides and
CC   the diglucuronide.
CC   The cDNAs contain identical 3' ends (1469 bp in length) to each
CC   other and to that of the human phenol transferase cDNA, HLUGP1
CC   (Harding et al., Proc. Natl. Aca. Sci. USA 85:8281 (1988)).
CC   In contrast, they have unique 5' ends.
XX
SQ   Sequence 2368 BP; 609 A; 532 C; 566 G; 661 T; 0 other;


  Query Match            68.0%;  Score 1576.8;  DB 13;  Length 2368;
  Best Local Similarity  82.3%;  Pred. No. 0;
  Matches 1847;  Conservative    0;  Mismatches  392;  Indels    5;  Gaps    3;

Qy     81 ctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcaccat 140
          ||  || |||   || ||| |  |||| |||||||  ||||| |||| |||| ||| |||
Db    101 ctgggctgagagtggaaaggtgttggtggtgcccactgatggcagcccctggctcagcat 160

Qy    141 gaggtcggtggtggagaaactcattctcaggggcatgaggtggttgtagtcatgccaga 200
          | ||  ||   ||  | | |||  |  ||| || ||  ||| ||| ||  |||  |||
Db    161 gcgggaggccttgcgggagctccatgccagaggccaccaggcggtggtcctcaccccaag 220

Qy    201 ggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttcata 260
          |    |   ||   |  |        || |||  |||   | ||| |   | |||
Db    221 aggtgaatatgcacatcaagaagagaaatttttcaccctgacagcctatgctgttccatg 280

Qy    261 taccctggaggatctggaccgggagttcaaggcttttgcccatgctcaat---ggaaagc 317
          ||||  | ||||  |  || || |       ||  |   | || |      |     | ||
Db    281 gacccagaaggaatttgatcgcgttacgctgggctacactcaagggttctttgaaacaga 340

Qy    318 acaagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttatt 377
          |||  | |  |  | |||||||  |   ||||        ||  |  | | |       |
Db    341 acatcttctgaagagatattctagaagtatggcaattatgaacaatgtatctttggccct 400

Qy    378 tttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagag 437
          |  |     |||   |    |  |  |  |  |     |  |     || | || |  |
Db    401 tcataggtgttgtgtggagctactgcataatgaggccctgatcaggcacctgaatgctac 460

Qy    438 ttcttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaata 497
          ||| |||||||  || ||    || ||  || |   ||| ||      |  | || || ||
Db    461 ttcctttgatgtggttttaacagaccccgttaacctctgcggggcggtgctggctaagta 520

Qy    498 tttctccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaagg 557
            | ||  | ||  | ||| | ||    |||   || |  ||  |||        | ||
Db    521 cctgtcgattcctgctgtgttttttttggaggtacattccatgtgacttagactttaaggg 580

Qy    558 tgcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgc 617
            |||||||  ||   ||||   |||||| | || |   | ||          |||||
Db    581 cacacagtgtccaaatccttcctcctatattcctaagttactaacgaccaattcagacca 640

Qy    618 catgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatg 677
          ||||||  |||  | | || ||   ||||    ||   |   |||        |   | ||
Db    641 catgacattcctgcaaagggtcaagaacatgctctaccctctggccctgtcctacatttg 700

Qy    678 ccaccgtttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttac 737
          |||   |||||       |     |    | ||||||||  || | || | |   || ||
Db    701 ccatacttttttctgccccttatgcaagtcttgcctctgagctttttcagagagaggtgac 760

Qy    738 ggagtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttgga 797
          |   | || || |  |||   ||||  | ||| ||||  |||  ||||||||   |||
Db    761 tgtccaggacctattgagctctgcatctgtctggctgtttagaagtgactttgtgaagga 820

Qy    798 ctatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggg 857
          || || |  ||| | |||||||| ||| ||||||||| || ||||||||    ||  |
Db    821 ttaccctaggcccatcatgcccaatatggtcttcattgggggcatcaactgttgccaacg 880
```

```
PT   Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT   for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT   types I and II
XX
PS   Disclosure; Fig 9A-I; 99pp; English.
XX
CC   Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
CC   been isolated.  They are referred to as HUGBr1 (AAQ27369) and HUGBr2
CC   (AAQ27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
CC   upon expression individually in COS-1 cells, encode isoforms that
CC   catalyse the formation of the two bilirubin monoglucuronides and
CC   the diglucuronide.
CC   The cDNAs contain identical 3' ends (1469 bp in length) to each
CC   other and to that of the human phenol transferase cDNA, HLUGP1
CC   (Harding et al., Proc. Natl. Aca. Sci. USA 85:8281 (1988)).
CC   In contrast, they have unique 5' ends.
XX
SQ   Sequence 2351 BP; 602 A; 540 C; 556 G; 653 T; 0 other;
```

```
  Query Match            68.2%;  Score 1581.4;  DB 13;  Length 2351;
  Best Local Similarity  82.5%;  Pred. No. 0;
  Matches 1861;  Conservative    0;  Mismatches  391;  Indels    5;  Gaps    4;

Qy      68 tgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatgggagcc 127
           |||||  |   |   |  || | || |||||| ||||| |  | ||  ||||||| ||||
Db      71 tgctgggcccagtggtgtcccatgctgggaagatactgttgatcccagtggatggcagcc 130

Qy     128 actggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgaggtggttg 187
           ||||| | | ||||     ||   |  || | ||    |  ||||| |||||  | ||||
Db     131 actggctgagcatgcttggggccatccagcagctgcagcagaggggacatgaaatagttg 190

Qy     188 tagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaagactt 247
           |  |    || || |     | | |  |  ||     |     || |||  ||||||| |
Db     191 tcctagcacctgacgcctcgttgtacatcagagacggagcattttacaccttgaagacgt 250

Qy     248 attcaacttcatataccctggaggatctggaccgggagtt--caaggcttttgcccatgc 305
           |  |     |||       |||||| || |   |   ||    || ||   ||  |
Db     251 accctgtgccattccaaagggaggatgtgaaagagtcttttgttagtctcgggcataatg 310

Qy     306 tcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgacat 365
           |   ||  ||  |    | |         | |  || |   ||      | | ||  |
Db     311 tttttgagaatgattctttcctgcagcgtgtgatcaaaacatacaagaaaataaaaaagg 370

Qy     366 ttttgact-tatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaat 424
             |   || |  |||| ||    ||       ||  |  |  |||| |  |  | |  |
Db     371 actctgctatgcttttgtctggctgttcccacttactgcacaacaaggagctcatggcct 430

Qy     425 acttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggcttaa 484
            | |    || ||    |||||||   || |    || ||||   |   ||  ||   |
Db     431 ccctggcagaaagcagctttgatgtcatgctgacggacccttccttccttgcagcccca 490

Qy     485 ttgttgccaaatatttctccctccccctccgtggtcttcgccaggggaatactttgccact 544
           | || ||| | ||  | || || ||| | ||  |||||      | | | |  |||  |
Db     491 tcgtggcccagtacctgtctctgcccactgtattcttcttgcatgcactgccatgcagcc 550

Qy     545 atcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttag 604
                 || | | ||||||||    ||  | |||||| || |||||   |||||
Db     551 tggaatttgaggctacccagtgcccccaacccattctcctacgtgcccaggcctctctcct 610

Qy     605 ggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggagg 664
           |||||| |||||| |||  | || | ||   ||||    ||||   |||
Db     611 ctcattcagatcacatgaccttcctgcagcgggtgaagaacatgctcattgccttttcac 670

Qy     665 aacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaattctcc 724
           |  | ||  | ||| ||    |||              |    | ||||| ||| | || |
Db     671 agaactttctgtgcgacgtggtttattccccgtatgcaacccttgcctcagaattccttc 730

Qy     725 aaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacgg 784
           | | |   || || |   | || || |  |||    ||||  | ||| ||||  |||  |
Db     731 agagagaggtgactgtccaggacctattgagctctgcatctgtctggctgtttagaagtg 790

Qy     785 actttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatca 844
```

```
                ||||||||   ||| || || |  ||| | |||||||| ||| | ||  |||||||| ||||
Db     791 actttgtgaaggattaccctaggcccatcatgcccaatatggtttttgttggtggaatca 850

Qy     845 actgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctggag 904
           ||||||  |||   |||  ||  |  |  |||||||||||||||||||||||||||||||
Db     851 actgccttcaccaaaatccactatcccaggaatttgaagcctacattaatgcttctggag 910

Qy     905 aacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaagaaag 964
           ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     911 aacatggaattgtggttttctctttggaatcaatggtctcagaaattccagagaagaaag 970

Qy     965 ctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactg 1024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     971 ctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactg 1030

Qy    1025 gaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacg 1084
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1031 gaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacg 1090

Qy    1085 atctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgttt 1144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1091 atctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgttt 1150

Qy    1145 atgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatgg 1204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1151 atgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatgg 1210

Qy    1205 acaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatga 1264
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1211 acaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatga 1270

Qy    1265 cttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggaga 1324
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    1271 cttctgaagatttagaaaatgctcaaaaagcagtcatcaatgacaaaagttacaaggaga 1330

Qy    1325 acatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccg 1384
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1331 acatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccg 1390

Qy    1385 tgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagccc 1444
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1391 tgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagccc 1450

Qy    1445 acgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcg 1504
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1451 acgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtgc 1510

Qy    1505 tgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttgg 1564
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1511 tgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttgg 1570

Qy    1565 ggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtggga 1624
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1571 ggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtggga 1630

Qy    1625 aataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtg 1684
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1631 aataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtg 1690

Qy    1685 ttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtg 1744
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    1691 ttaaattcattttattcttattaaggaaatactttgcat-aattaatcagccccagagtg 1749

Qy    1745 ctttaaaaaattctctttaaataaaaataatagactcgctagtcagtaaagatatttgaat 1804
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1750 ctttaaaaaattctctttaaataaaaataatagactcgctagtcagtaaagatatttgaat 1809

Qy    1805 atgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggac 1864
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1810 atgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggac 1869

Qy    1865 gtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttggga 1924
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy    781 acggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggt 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 acggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggt 840

Qy    841 atcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttct 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 atcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttct 900

Qy    901 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 960

Qy    961 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1020

Qy   1021 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaa 1080

Qy   1081 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1140

Qy   1141 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1200

Qy   1201 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1260

Qy   1261 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1320

Qy   1321 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1380

Qy   1381 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1440

Qy   1441 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1500

Qy   1501 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1560

Qy   1561 ttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggt 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggt 1620

Qy   1621 gggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 gggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatc 1680

Qy   1681 agtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccag 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 agtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccag 1740

Qy   1741 agtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatattt 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 agtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatattt 1800

Qy   1801 gaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcaga 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 gaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcaga 1860
```

```
Qy   1861 ggacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgttt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 ggacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgttt 1920

Qy   1921 gggagtgccggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaat 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 gggagtgccggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaat 1980

Qy   1981 cttatctttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcct 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 cttatctttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcct 2040

Qy   2041 catctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacc 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 catctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacc 2100

Qy   2101 tttagagtctaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgattt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 tttagagtctaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgattt 2160

Qy   2161 cttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagata 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 cttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagata 2220

Qy   2221 atcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagt 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 atcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagt 2280

Qy   2281 gctgggcaagtttacttttttttctgatgtttcctacaact 2320
          ||||||||||||||||||||||||||||||||||||||||
Db   2281 gctgggcaagtttacttttttttctgatgtttcctacaact 2320
```

```
RESULT   3
AAQ27369
ID   AAQ27369 standard; cDNA; 2351 BP.
XX
AC   AAQ27369;
XX
DT   27-JAN-1993  (first entry)
XX
DE   HUG-Br1.
XX
KW   Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
KW   monoglucuronide; diglucuronide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             16..784
FT                   /*tag=  a
FT   polyA_signal    2330..2335
FT                   /*tag=  b
FT                   /number= 1
FT   polyA_signal    2338..2343
FT                   /*tag=  c
FT                   /number= 2
XX
PN   WO9212987-A.
XX
PD   06-AUG-1992.
XX
PF   10-JAN-1992;   92WO-US00282.
XX
PR   10-JAN-1991;   91US-0639453.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Owens IS,  Ritter JK;
XX
DR   WPI; 1992-284593/34.
DR   P-PSDB; AAR26153.
XX
```

```
Qy   1801 gaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcaga 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 gaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcaga 1860

Qy   1861 ggacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgttt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 ggacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgttt 1920

Qy   1921 gggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaat 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 gggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaat 1980

Qy   1981 cttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcct 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 cttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcct 2040

Qy   2041 catctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacc 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 catctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacc 2100

Qy   2101 tttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgattt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 tttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgattt 2160

Qy   2161 cttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagata 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 cttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagata 2220

Qy   2221 atcattgcttatgtcaaatggagctgaatttgataaaaaacccaaaatacagctatgaagt 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 atcattgcttatgtcaaatggagctgaatttgataaaaaacccaaaatacagctatgaagt 2280

Qy   2281 gctgggcaagtttactttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||||||
Db   2281 gctgggcaagtttactttttttctgatgtttcctacaact 2320
```

```
RESULT   2
ABA05199
ID   ABA05199 standard; cDNA; 2320 BP.
XX
AC   ABA05199;
XX
DT   04-MAR-2002  (first entry)
XX
DE   Human uridine 5'diphosphate glucuronyltransferase UGT1A9 cDNA #2.
XX
KW   Human; uridine 5'diphosphate glucuronyltransferase enzyme; UGT1A9;
KW   flavopiridol; cancer; gastrointestinal disease; parasitic infection;
KW   cytostatic; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             19..1611
FT                   /*tag= a
FT                   /product= "UGT1A9"
XX
PN   WO200180896-A2.
XX
PD   01-NOV-2001.
XX
PF   12-APR-2001; 2001WO-US12526.
XX
PR   21-APR-2000; 2000US-0553829.
XX
PA   (ARCH-) ARCH DEV CORP.
XX
PI   Ratain MJ,  Innocenti F,  Iyer L;
XX
DR   WPI; 2002-075093/10.
DR   P-PSDB; ABB04430.
XX
PT   Combinations of flavopiridol and an agent that increases conjugative
PT   enzyme activity or glucuronosyltransferase activity, with reduced side
PT   effects, for treating cancer -
XX
PS   Claim 81; Page 141-144; 145pp; English.
XX
CC   The present invention relates to a method of reducing the toxicity of
CC   flavopiridol by administration in combination with a second agent that
CC   increases conjugative enzyme activity or glucuronosyltransferase
CC   activity. This second agent should be capable of inhibiting biliary
CC   transport and may be a uridine 5'diphosphate glucuronyltransferase such
CC   as that encoded by the present sequence (designated UGT1A9). The method
CC   can be used in the treatment of cancer, gastrointestinal diseases and
CC   parasitic diseases.
XX
SQ   Sequence 2320 BP; 608 A; 495 C; 547 G; 670 T; 0 other;
```

```
  Query Match           100.0%;  Score 2320;  DB 24;  Length 2320;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2320;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60

Qy     61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120

Qy    121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgag 180

Qy    181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240

Qy    241 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 300

Qy    301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360

Qy    361 gacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagta 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagta 420

Qy    421 gaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 gaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggc 480

Qy    481 ttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgc 540

Qy    541 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 600

Qy    601 ttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttg 660

Qy    661 gaggaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaatt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 gaggaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaatt 720

Qy    721 ctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcga 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcga 780
```

XX
DR WPI; 2002-075093/10.
DR P-PSDB; ABB04429.
XX
PT Combinations of flavopiridol and an agent that increases conjugative
PT enzyme activity or glucuronosyltransferase activity, with reduced side
PT effects, for treating cancer -
XX
PS Claim 80; Page 137-140; 145pp; English.
XX
CC The present invention relates to a method of reducing the toxicity of
CC flavopiridol by administration in combination with a second agent that
CC increases conjugative enzyme activity or glucuronosyltransferase
CC activity. This second agent should be capable of inhibiting biliary
CC transport and may be a uridine 5'diphosphate glucuronyltransferase such
CC as that encoded by the present sequence (designated UGT1A9). The method
CC can be used in the treatment of cancer, gastrointestinal diseases and
CC parasitic diseases.
XX
SQ Sequence 2320 BP; 608 A; 495 C; 547 G; 670 T; 0 other;

Query Match 100.0%; Score 2320; DB 24; Length 2320;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2320, Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60

Qy      61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120

Qy     121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgag 180
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgag 180

Qy     181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240

Qy     241 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 300

Qy     301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360

Qy     361 gacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagta 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 gacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagta 420

Qy     421 gaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 gaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggc 480

Qy     481 ttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgc 540

Qy     541 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 600

Qy     601 ttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttg 660

Qy     661 gaggaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaatt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 gaggaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaatt 720

Qy     721 ctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcga 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcga 780

Qy     781 acggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggt 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 acggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggt 840

Qy     841 atcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttct 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 atcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttct 900

Qy     901 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 960

Qy     961 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1020

Qy    1021 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaa 1080

Qy    1081 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1140

Qy    1141 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1200

Qy    1201 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1260

Qy    1261 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1320

Qy    1321 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1380

Qy    1381 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1440

Qy    1441 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1500

Qy    1501 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1560

Qy    1561 ttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggt 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 ttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggt 1620

Qy    1621 gggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 gggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatc 1680

Qy    1681 agtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccag 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 agtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccag 1740

Qy    1741 agtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatattt 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 agtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatattt 1800
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 21, 2002, 06:34:42 ; Search time 682.45 Seconds
(without alignments)
5836.679 Million cell updates/sec

Title: US-09-835-082-1
Perfect score: 2320
Sequence: 1 ctcagctgcagttctctgat.....ttctgatgtttcctacaact 2320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2320	100.0	2320	24	ABA05198	Human uridine 5'di
	2	2320	100.0	2320	24	ABA05199	Human uridine 5'di
	3	1581.4	68.2	2351	13	AAQ27369	HUG-Br1. Homo sap
	4	1576.8	68.0	2368	13	AAQ27370	HUG-Br2. Homo sap
	5	1551.8	66.9	1655	23	AAS91252	DNA encoding novel
	6	1004.4	43.3	18887	24	AAS18543	DNA encoding UDP g
	7	998.6	43.0	1001	21	AAH51403	Human UGT1A7 relat
	8	870.6	37.5	1001	21	AAH51404	Human UGT1A7 relat
	9	798	34.4	1667	13	AAQ33027	Exon 5 from the UG
	10	769.6	33.2	1001	21	AAH51405	Human UGT1A7 relat
	11	766.6	33.0	930	21	AAZ45116	UDP-glucuronosyltr
	12	741.2	31.9	951	21	AAZ45115	UDP-glucuronosyltr
	13	735	31.7	735	21	AAZ45118	UDP-glucuronosyltr
	14	659.8	28.4	759	21	AAZ45117	UDP-glucuronosyltr
	15	578.8	24.9	874	24	ABL01468	Murine apoptosis r
	16	364.2	15.7	391	20	AAZ07208	Human lung tumour
	17	364.2	15.7	391	21	AAC79066	Human lung tumour-
	18	364.2	15.7	391	23	AAD23141	Human lung tumour-
	19	359	15.5	1001	21	AAH51406	Human UGT1A7 relat
	20	353	15.2	1001	21	AAH51407	Human UGT1A7 relat
	21	345.2	14.9	1854	21	AAZ95200	Human UDP-glucuron
	22	345	14.9	1766	23	AAS91380	DNA encoding novel
	23	344.4	14.8	2107	19	AAV15900	Uridine diphospho-
	24	342	14.7	380	21	AAC98723	Human colon cancer
	25	339.8	14.6	2092	21	AAZ95199	Human UDP-glucuron
	26	333.2	14.4	1976	21	AAZ95206	Human UDP-glucuron
	27	332.8	14.3	1650	21	AAC65396	Human carbohydrate
	28	327.2	14.1	2802	23	AAS69711	DNA encoding novel
	29	325.6	14.0	2111	23	AAS69712	DNA encoding novel
	30	322.4	13.9	1636	24	AAD24666	Human drug metabol
	31	322.4	13.9	1722	22	AAD06860	Human breast cance
	32	322.4	13.9	2974	22	AAS46185	Human DNA encoding
	33	318	13.7	1859	23	AAS69710	DNA encoding novel
	34	314.6	13.6	1713	22	AAD06844	Human breast cance
	35	302.4	13.0	508	21	AAC98478	Human colon cancer
	36	296	12.8	1001	21	AAH51408	Human UGT1A7 relat
	37	282	12.2	2025	23	AAS91253	DNA encoding novel
	38	230.4	9.9	366	21	AAA87085	Rat hepatocyte car
	39	222.4	9.6	775	13	AAQ33026	Exons 3-4 from the
	40	220.8	9.5	273	21	AAA87491	Rat hepatocyte car
c	41	203	8.8	2448	23	AAS72022	DNA encoding novel
c	42	203	8.8	2448	23	AAS74393	DNA encoding novel
c	43	203	8.8	2448	23	AAS78611	DNA encoding novel
c	44	203	8.8	2448	23	AAS78773	DNA encoding novel
c	45	203	8.8	2448	23	AAS84057	DNA encoding novel

ALIGNMENTS

RESULT 1
ABA05198
ID ABA05198 standard; cDNA; 2320 BP.
XX
AC ABA05198;
XX
DT 04-MAR-2002 (first entry)
XX
DE Human uridine 5'diphosphate glucuronyltransferase UGT1A9 cDNA #1.
XX
KW Human; uridine 5'diphosphate glucuronyltransferase enzyme; UGT1A9;
KW flavopiridol; cancer; gastrointestinal disease; parasitic infection;
KW cytostatic; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 19..1611
FT /*tag= a
FT /product= "UGT1A9"
XX
PN WO200180896-A2.
XX
PD 01-NOV-2001.
XX
PF 12-APR-2001; 2001WO-US12526.
XX
PR 21-APR-2000; 2000US-0553829.
XX
PA (ARCH-) ARCH DEV CORP.
XX
PI Ratain MJ, Innocenti F, Iyer L;

```
;    APPLICANT:  BELKNAP, WILLIAM R
;    TITLE OF INVENTION:  DNA SEQUENCES ENCODING SOLANIDINE
;    TITLE OF INVENTION:  UDP-GLUCOSE GLUCOSYLTRANSFERASE AND USE TO REDUCE
;    TITLE OF INVENTION:  GLYCOALKALOIDS IN SOLANACEOUS PLANTS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NANCY J. PARSONS
;      STREET:  800 BUCHANAN ST.
;      CITY:  ALBANY
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94710
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/797,226
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PARSONS, NANCY J
;      REGISTRATION NUMBER:  40,364
;      REFERENCE/DOCKET NUMBER:  0011.97
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 559-5731
;      TELEFAX:  (510) 559-5777
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1607 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Solanum tuberosum
;      STRAIN:  cv. Lemhi Russet
;      INDIVIDUAL ISOLATE:  SGT1750
;      DEVELOPMENTAL STAGE:  mature, somatic
;      TISSUE TYPE:  tuber
;    IMMEDIATE SOURCE:
;      LIBRARY:  lambda gtll cDNA library
;      CLONE:  SGT 1750
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20..1486
;      OTHER INFORMATION:  /product= "solanidine
;      OTHER INFORMATION:  glucosyltransferase"
;    PUBLICATION INFORMATION:
;      AUTHORS:  Moehs, Charles P.
;      AUTHORS:  Allen, Paul V.
;      AUTHORS:  Friedman, Mendel
;      AUTHORS:  Belknap, William R.
;      TITLE:  Cloning and expression of solanidine
;      TITLE:  UDP-glucose glucosyltransferase from potato
;      JOURNAL:  The Plant Journal
;      VOLUME:  11
;      ISSUE:  2
;      DATE:  1997
;      RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 488
US-08-797-226-1


  Query Match             1.9%;  Score 44.8;  DB 2;  Length 1607;
  Best Local Similarity   53.4%;  Pred. No. 0.0072;
  Matches   94;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy   1068 gtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcctttatcacccatgc 1127
          |||| | || ||         | | |  ||  |    ||  | |  || || || |||
Db   1069 GTGGGTCCCACAGCTTACGATCATGGAACATTCAGCAACAGGCGGGTTCATGACTCATTG 1128

Qy   1128 tggttcccatggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgccctt 1187
          |||| |  ||   |||   |||  |||      |||||| || ||| | |    |||  |
Db   1129 TGGTACTAATTCGGTTCTGGAAGCCATCACTTTTGGCGTGCCAATGATAACATGGCCACT 1188

Qy   1188 gtttggtgatcagatggacaatgcaaagcgcatggagactaagggagctggagtga 1243
           | || ||||||  |  |||| |  |||    | |||  || ||||   ||| | |
Db   1189 TTATGCTGATCAATTCTACAACGAGAAGGTAGTCGAGGTTAGGGGATTGGGAATCA 1244


RESULT  15
PCT-US92-00282-37
; Sequence 37, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  51 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-37


  Query Match             1.9%;  Score 43.6;  DB 5;  Length 51;
  Best Local Similarity   92.0%;  Pred. No. 0.0021;
  Matches   46;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy   1444 cacgacctcacctggtaccagtaccattccttggacgtgattggtttcct 1493
           |||||||||||||||| ||||||||| || ||||| ||||||||||||||
Db      1 CACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGTTTCCT 50


Search completed: August 21, 2002, 11:59:18
Job time: 32806 sec
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  513 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-20


  Query Match             5.8%;  Score 135;  DB 5;  Length 513;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-29;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     873 ggaatttgaagcctacattaatgcttctggagaacatggaattgtggttttctctttggg 932
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     230 GGAATTTGAAGCCTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTCTCTTTGGG 289

Qy     933 atcaatggtctcagaaattccagagaagaaagctatggcaattgctgatgctttgggcaa 992
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     290 ATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATGGCAATTGCTGATGCTTTGGGCAA 349

Qy     993 aatccctcagacagt 1007
           |||||||||||||||
Db     350 AATCCCTCAGACAGT 364


RESULT  13
US-08-522-421-1
; Sequence 1, Application US/08522421
; Patent No. 5908973
;  GENERAL INFORMATION:
;    APPLICANT:  Abu-Bakar, Umi Kalsom
;    APPLICANT:  Barton, Sarah Louise
;    APPLICANT:  Gallego-Veigas, Pedro Pablo
;    APPLICANT:  Gray, Julie Elizabeth
;    APPLICANT:  Grierson, Donald
;    APPLICANT:  Lowe, Alexandra Louise
;    APPLICANT:  Picton, Steve
;    APPLICANT:  Whotton, Lee Colin
;    TITLE OF INVENTION:  DNA, DNA CONSTRUCTS, CELLS AND PLANTS
;    TITLE OF INVENTION:  DERIVED THEREFROM
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pillsbury Madison & Sutro, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/522,421
;      FILING DATE:  11-JAN-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305868.3
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305869.1
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305859.2
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305865.9
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305866.7
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305867.5
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305860.0
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305862.6
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9314351.9
;      FILING DATE:  12-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9320988.0
;      FILING DATE:  12-OCT-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1669 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  ERT1B
US-08-522-421-1


  Query Match             2.4%;  Score 54.6;  DB 2;  Length 1669;
  Best Local Similarity   59.2%;  Pred. No. 1.1e-05;
  Matches   93;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy    1059 acttgttaagtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcctttat 1118
           | | || || |||  ||| |||||  |  ||||||  || ||   |      |  ||| |
Db     965 AATAGTGAAATGGGCACCACAAAAACAGGTGCTTGCACATCCCGCGGTAGCAGGGTTTTT 1024

Qy    1119 cacccatgctggttcccatggtgtttatgaaagcatatgcaatggcgttcccatggtgat 1178
           |||  |||  |||||   ||  |     ||||||  |||||  | |  || || ||||||
Db    1025 CACTCATTGTGGTTGGAATTCTACGCTTGAAAGTATATGTGAAGAAGTCCCTATGGTGTG 1084

Qy    1179 gatgcccttgtttggtgatcagatggacaatgcaaag 1215
           | ||| ||  | |  || ||  |||  || |||| |
Db    1085 CAGGCCATTTCTAGCAGACCAACTGGTGAACGCAAGG 1121


RESULT  14
US-08-797-226-1
; Sequence 1, Application US/08797226
; Patent No. 5959180
;  GENERAL INFORMATION:
;    APPLICANT:  MOEHS, CHARLES P
;    APPLICANT:  ALLEN, PAUL V
;    APPLICANT:  ROCKHOLD, DAVID R
;    APPLICANT:  STAPLETON, ANDREW
;    APPLICANT:  GARBARINO, JOAN E
;    APPLICANT:  FRIEDMAN, MENDEL
```

```
Qy    560 cacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgcca 619
          ||||||  |   ||||  |||||| | || |   | ||          |||||   ||
Db    784 TACAGTG-TCAAATCCTTCCTCCTATATTCCTAAGTTACTAACGACCAATTCAGACCACA 842

Qy    620 tgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatgcc 679
          |||| |||  | | || ||   ||||    ||   |   |||         |   | ||||
Db    843 TGACATTCCTGCAAAGGGTCAAGAACATGCTCTACCCTCTGGCCCTGTCCTACATTTGCC 902

Qy    680 accgttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttacgg 739
          |   |||||       |     |    | ||||||||  || | || | |   ||  | |
Db    903 ATACTTTTTCTGCCCCTTATGCAAGTCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTCAG 962

Qy    740 agtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttggact 799
           |   |||||   |||| |  ||||  | ||| |||| |||  |||||||||  ||||||
Db    963 TGGTGGATCTTGTCAGCTATGCATCCGTGTGGCTGTTCCGAGGGGACTTTGTGATGGACT 1022

Qy    800 atcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcagggaa 859
          | ||||  ||  | ||||||||||||| ||||||||| || ||||||||     | || |
Db   1023 ACCCCAGGCCGATCATGCCCAACATGGTCTTCATTGGGGGCATCAACTGTGCCAACGGGA 1082

Qy    860 agccgttgcctatggaattt 879
          ||||  |  ||  ||  | |
Db   1083 AGCCACTATCTCAGGTCTGT 1102


RESULT  11
PCT-US92-00282-8
; Sequence 8, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1219 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-8


  Query Match             6.2%;  Score 143.6;  DB 5;  Length 1219;
  Best Local Similarity   50.8%;  Pred. No. 2.3e-31;
  Matches  419;  Conservative    0;  Mismatches  399;  Indels    6;  Gaps    3;

Qy     47 ttcctctatgtgtgtgtctgctgacctgtggctttgccgaggcagggaagctactgg 106
          || ||  | | || |   | |        || ||| | |  |  |  | ||||| ||||
Db    245 TTTCTGCAGGGGTTTTCTTCTTAGCACTTTGGGGCATGGTTGTAGGTGACAAGCTGCTGG 304

Qy    107 tagtgcccatggatgggagccactggttcaccatgaggtcggtggtggagaaactcattc 166
          | || ||   ||| || |||||||||| | |  |||| |    | || |||   |||| |
Db    305 TGGTCCCTCAGGACGGAAGCCACTGGCTTAGTATGAAGGATATAGTTGAGGTTCTCAGTG 364

Qy    167 tcaggggggcatgaggtggttgtagtcatgccagaggtgagttggcaactgggaagatcac 226
          | ||||  ||||| | || || ||  |||| || || | || ||   ||  |  |||
Db    365 ACCGGGGTCATGAGATTGTAGTGGTGGTGCCTGAAGTTAATTTGCTTTTGAAAGAATCCA 424

Qy    227 tgaattgcacagtgaagacttattcaacttcatataccctggaggatctggaccgggagt 286
          | | ||||   || |  ||| ||    | |||  ||  || || ||| |        |
Db    425 AATACTACACAAGAAAAATCTATCCAGTGCCGTATGACCAAGAAGAGCTGAAGAACCGTT 484

Qy    287 tcaaggcttttgcccatgctcaatggaaagcacaagtacgaagtatatattctctattaa 346
          | |  | ||||    |  ||| |     |  | |  |     |    | |||    |
Db    485 ACCAATCATTTGGAAACAATCACTTTGCTGAGCGATCATTCCTAACTGCTCCTCAGACAG 544

Qy    347 tgggttcatacaatgacatt----tttgacttatttttttcaaattgcaggagtttgttt 402
          |      | ||    |||    || | || ||  ||    || |||  |||  |  |
Db    545 AGTACAGGAATAACATGATTGTTATTGGCCTGATACTTCATCAACTGCCAGAGCCTCCTG 604

Qy    403 aaagacaaaaaattagtagaatacttaaaggagagttctttttgatgcagtgtttctcgat 462
          | ||||   |     |  | | ||| ||||||||    ||||||||  | ||    ||
Db    605 CAGGACAGCGACACCCTGAACTTCTTTAAGGAGAGCAAGTTTGATGCTCTTTTCACAGAC 664

Qy    463 ccttttgataactgtggcttaattgttgccaaatatttctccctccccctccgtggtcttc 522
          ||         ||||||  | ||  | || |    ||   ||  ||          ||
Db    665 CCAGCCTTACCCTGTGGGGTGATCCTGGCGAGTATTTGGGCCTACCATCTGTGTACCTCT 724

Qy    523 gccaggggaatactttgccactatcttgaagaaggtgcacagtgccctgctcctctttcc 582
          |  |||   | | |   | ||       | |    ||| |   ||   | |    |  |
Db    725 TCAGGGGTTTTCCGTGTTCCCTGGAGCATACATTCAGCAGAAGCCCAGACCCTGTGTCTC 784

Qy    583 tatgtccccagaattctcttagggttctcagatgccatgactttcaaggagagagtacgg 642
             |||||||      |  |  ||| |||||   |||||||||     |  ||||
Db    785 ACATTCCCCAGGTGCTACACAAAGTTTTCAGACCACATGACTTTTTCCCAACGAGTGGCC 844

Qy    643 aaccacatcatgcacttggaggaacatttattatgccaccgttttttcaaaaatgcccta 702
          |||  | |  |  | |||  ||| |  |   |||   |  || | ||  |||      |
Db    845 AACTTCCTTGTTAATTTGTTGGAGCCCTATCTATTTTATTGTCTGTTTTCAAAGTATGAA 904

Qy    703 gaaatagcctctgaaattctccaaacacctgtt-acggagtatgatctctacagccacac 761
          ||| | || || |   | ||| | | |  |||  |   | |    |      ||   |
Db    905 GAACTCGCATCAGCTGTCCTCAAGAGAGATGTGGATATAATCACCTTATATCATGAAGGT 964

Qy    762 atcaatttggttgttgcgaacggactttgttttggactatcccaaacccgtgatgcccaa 821
          ||  ||||| ||||  ||   |||||||  | || ||||| |  || || ||||||||
Db    965 CTCTGTTTCGCTGTTAAGATATGACTTTGTGCTTGAATATCCTAGGCCGGTCATGCCCAA 1024

Qy    822 catgatcttcattggtggtatcaactgccatcagggaaagccgt 865
          | || |||||||||| ||||||||||  |  || | ||| | |
Db   1025 C-TGGTCTTCATTGGAGGTATCAACTGTAAGAAGAGGAAGACTT 1067


RESULT  12
PCT-US92-00282-20
; Sequence 20, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
```

```
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-18


  Query Match             7.0%;  Score 161.8;  DB 5;  Length 1190;
  Best Local Similarity   51.8%;  Pred. No. 1.3e-36;
  Matches  414;  Conservative    0;  Mismatches  382;  Indels    3;  Gaps    2;

Qy      68 tgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatgggagcc 127
           |||||  |   |   |  || | || |||||| ||||| |  | ||  ||||||| ||||
Db     140 TGCTGGGCCCAGTGGTGTCCCATGCTGGGAAGATACTGTTGATCCCAGTGGATGGCAGCC 199

Qy     128 actggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgaggtggttg 187
           ||||| | | ||||     ||   |  || | ||    |  ||||| |||||  | ||||
Db     200 ACTGGCTGAGCATGCTTGGGGCCATCCAGCAGCTGCAGCAGAGGGGACATGAAATAGTTG 259

Qy     188 tagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaagactt 247
           |  |    || || |    | | |  |  ||     |    || |||  ||||||| |
Db     260 TCCTAGCACCTGACGCCTCGTTGTACATCAGAGACGGAGCATTTTACACCTTGAAGACGT 319

Qy     248 attcaacttcatataccctggaggatctggaccgggagtt--caaggcttttgcccatgc 305
           |  |     ||       ||||||| || |   |  ||    || ||   ||  |
Db     320 ACCCTGTGCCAATCCAAAGGGAGGATGTGAAAGAGTCTTTTGTTAGTCTCGGGCATAATG 379

Qy     306 tcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgacat 365
           |   ||  ||  |   | |        | |  || |  ||      | | ||  |
Db     380 TTTTTGAGAATGATTCTTTCCTGCAGCGTGTGATCAAAACATACAAGAAAATAAAAAAGG 439

Qy     366 ttttgact-tatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaat 424
           |   || |  |||| ||    ||       ||  |  |  ||||  |  |  | |  |
Db     440 ACTCTGCTATGCTTTTGTCTGGCTGTTCCCACTTACTGCACAACAAGGAGCTCATGGCCT 499

Qy     425 acttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggcttaa 484
           | |    || ||    |||||||   || |    || |||||   |   ||  ||   |
Db     500 CCCTGGCAGAAAGCAGCTTTGATGTCATGCTGACGGACCCTTTCCTTCCTTGCAGCCCCA 559

Qy     485 ttgttgccaaatatttctccctccccctccgtggtcttcgccaggggaatactttgccact 544
           | || ||| | ||  | || || ||| | ||  |||||      | | | |  |||  |
Db     560 TCGTGGCCCAGTACCTGTCTCTGCCCACTGTATTCTTCTTGCATGCACTGCCATGCAGCC 619

Qy     545 atcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttag 604
                 || | | |||||||||    ||  | |||||  || |||||   |||||
Db     620 TGGAATTTGAGGCTACCCAGTGCCCCAACCCATTCTCCTACGTGCCCAGGCCTCTCTCCT 679

Qy     605 ggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggagg 664
               ||||||  ||||||  |||  | ||  | ||   ||||    ||||   |||
Db     680 CTCATTCAGATCACATGACCTTCCTGCAGCGGGTGAAGAACATGCTCATTGCCTTTTCAC 739

Qy     665 aacatttattatgccaccgtttttttcaaaaatgccctagaaatagcctctgaaattctcc 724
           |  | ||  | ||| ||    |||              |   | |||||  ||| | || |
Db     740 AGAACTTTCTGTGCGACGTGGTTTATTCCCCGTATGCAACCCTTGCCTCAGAATTCCTTC 799

Qy     725 aaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacgg 784
           | | |   || || |   | || || |  |||    ||||  | ||| ||||  |||  |
Db     800 AGAGAGAGGTGACTGTCCAGGACCTATTGAGCTCTGCATCTGTCTGGCTGTTTAGAAGTG 859

Qy     785 actttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatca 844
           |||||||   |||  | || |  ||| | ||||||| ||| | ||  |||||| ||||
Db     860 ACTTTGTGAAGGATGACCCTAGGCCCATCATGCCCAATATGGTTTTTGTTGGTGGAATCA 919

Qy     845 actgccatcagggaaagcc 863
           |||||| |||   ||| ||
Db     920 ACTGCCTTCACCAAAATCC 938


RESULT  10
PCT-US92-00282-12
; Sequence 12, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1197 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-12


  Query Match             6.9%;  Score 161.2;  DB 5;  Length 1197;
  Best Local Similarity   52.6%;  Pred. No. 2e-36;
  Matches  421;  Conservative    0;  Mismatches  373;  Indels    6;  Gaps    3;

Qy      85 gccgaggcagggaagctactggtagtgcccatggatgggagccactgg--ttcaccatga 142
           || |||   || ||| |  |||| |||||||  ||||| |||| ||||  |    ||||
Db     304 GCTGAGAGTGGAAAGGTGTTGGTGGTGCCCACTGATGGCAGCCCCTGGTCTCAGTCATGC 363

Qy     143 ggtcggtggtggagaaactcattctcaggggcatgaggtggttgtagtcatgccagagg 202
           ||  ||   ||  | | |||  |  ||| || ||  ||| ||| ||  |||  |||||||
Db     364 GGGAGGCCTTGCGGGAGCTCCATGCCAGAGGCCACCAGGCGGTGGTCCTCACCCCAGAGG 423

Qy     203 tgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttcatata 262
           ||| |  |||  |   |  |         || |||  |||   | ||| |   | |||  |
Db     424 TGAATATGCACATCAAAGAAGAGAAATTTTTCACCCTGACAGCCTATGCTGTTCCATGGA 483

Qy     263 ccctggaggatctggaccgggagttcaaggcttttgcccatgctcaat---ggaaagcac 319
           ||| | ||||  | || || |       ||  |   | || |      |     | || ||
Db     484 CCCAGAAGGAATTTGATCGCGTTACGCTGGGCTACACTCAAGGGTTCTTTGAAACAGAAC 543

Qy     320 aagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttatttt 379
           |  | |  |  | |||||||  |   ||||        ||  |  | | |       ||
Db     544 ATCTTCTGAAGAGATATTCTAGAAGTATGGCAATTATGAACAATGTATCTTTGGCCCTTC 603

Qy     380 tttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagagtt 439
           |     |||  |    |  | |  |  |  |      |  |     || | || |  | ||
Db     604 ATAGGTGTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGGCACCTGAATGCTACTT 663

Qy     440 cttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaatatt 499
           | |||||||  || ||    || ||  || |   ||||||      |  | || || ||
Db     664 CCTTTGATGTGGTTTTAACAGACCCCGTTAACCTCTGTGGGGCGGTGCTGGCTAAGTACC 723

Qy     500 tctccctccccctccgtggtcttcgccaggggaatactttgccactatcttgaagaaggtg 559
           | ||  | ||  | ||| | ||    |||   || |  ||  |||         | |
Db     724 TGTCGATTCCTGCTGTGTTTTTTTGGAGGTACATTCCATGTGACTTAGACTTTAAGGCCA 783
```

```
Qy    858 aaagccgttgcctatgg 874
          ||,||  |  ||  ||
Db    957 GAAGCCACTATCTCAGG 973


RESULT   8
PCT-US92-00282-14
; Sequence 14, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1066 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-14


  Query Match             7.9%;  Score 183.6;  DB 5;  Length 1066;
  Best Local Similarity   53.5%;  Pred. No. 6.7e-43;
  Matches  429;  Conservative    0;  Mismatches  369;  Indels    4;  Gaps    2;

Qy      81 ctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcaccat 140
           ||  || |||   || ||| |  |||| ||||||||  |||| |||||||||| ||| |||
Db     202 CTGGGCTGAGAGTGGAAAGGTGTTGGTGGTGCCCATTCATGGCAGCCACTGGCTCAGCAT 261

Qy     141 gaggtcggtggtggagaaactcattctcaggggcatgaggtggttgtagtcatgccaga 200
           | ||  |||  ||  | | |||  |  ||| || ||  |||  || ||  |||  |||||
Db     262 GCGGGAGGTCTTGCGGGAGCTCCATGCCAGAGGCCACCAGGCAGTGGTCCTCACCCCAGA 321

Qy     201 ggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttcata 260
           |||||  |  |||  |   |  |        || |||  |||  || ||| | | ||| |
Db     322 GGTGAATATGCACATCAAAGAAGAGAACTTTTTCACCCTGACAACCTATGCCATTTCGTG 381

Qy     261 taccctggaggatctggaccgggagttcaaggcttttgcccatgctcaat---ggaaagc 317
           ||||  ||| ||  | || ||  |  |   ||      | ||     | |     | ||
Db     382 GACCCAGGATGAATTTGATCGCCATGTGCTGGGCCACACTCAACTGTACTTTGAAACAGA 441

Qy     318 acaagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttatt 377
           |||  |  |  | || ||    |  ||||     |  ||  | || | |    | |
Db     442 ACATTTTCTGAAGAAATTTTTCAGAAGTATGGCAATGTTGAACAATATGTCTTTGGTCTA 501

Qy     378 tttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagag 437
           |  |    |||   |   |  |  |  |  |     |  |     || | || |  |
Db     502 TCATAGGTCTTGTGTGGAGCTACTACATAATGAGGCCCTGATCAGGCACCTGAATGCTAC 561

Qy     438 ttctttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaata 497
           |||  |||||||  || ||    || ||  || |   ||| |    | |  | || || ||
Db     562 TTCCTTTGATGTGGTTTTAACAGACCCCGTTAACCTCTGCGCGGCAGTGCTGGCTAAGTA 621

Qy     498 tttctcccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaagg 557
           |  ||  | ||  | ||| | ||    |||   || |  ||  | |         | ||
Db     622 CCTGTCGATTCCTACTGTGTTTTTTTTGAGGAACATTCCATGTGATTTAGACTTTAAGGG 681

Qy     558 tgcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgc 617
           |||||||  ||    |||   |||||| | || ||| | ||   |    |||||
Db     682 CACACAGTGTCCAAACCCTTCCTCCTATATTCCTAGATTACTAACAACCAATTCAGACCA 741

Qy     618 catgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatg 677
           ||||||  | || | | || ||   ||||    ||   |  |||          |  | ||
Db     742 CATGACATTCATGCAAAGGGTCAAGAACATGCTCTACCCTCTGGCCCTGTCCTACATTTG 801

Qy     678 ccaccgtttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttac 737
           |||   |||||         |          | ||||||||  || | || | |   || |
Db     802 CCATGCTTTTTTCTGCTCCTTATGCCAGCCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTC 861

Qy     738 ggagtatgatctctacagccacacatcaatttggttgttgcgaacggacttttgttttgga 797
           | |   ||| |  ||| ||  |||| | ||| |||| |||  |||||||||  ||||
Db     862 AGTGGTGGATATTCTCAGTCATGCATCTGTGTGGCTGTTCCGAGGGGACTTTGTGATGGA 921

Qy     798 ctatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggg 857
           |||  ||||  ||  | ||||||||||||  |||||||||| || ||||||||      | |
Db     922 CTACCCCAGGCCAATCATGCCCAACATGGTCTTCATTGG-GGCATCAACTGTGCCAACAG 980

Qy     858 aaagccgttgcctatggaattt 879
           |||||  |  ||  ||  | |
Db     981 GAAGCCACTATCTCAGGTCTGT 1002


RESULT   9
PCT-US92-00282-18
; Sequence 18, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1190 base pairs
```

```
Db    303 GGTGAATATGTACATCAAAGAAGAGAACTTTTTCACCCTGACAACGTATGCCATTTCATG 362

Qy    261 taccctggaggatctggaccgggagttcaaggcttttgcccatgctcaat---ggaaagc 317
          |||| ||| || | || ||   ||  || |   | ||  |    |     | ||
Db    363 GACCCAGGACGAATTTGATCGCCTTTTGCTGGGTCACACTCAATCGTTCTTTGAAACAGA 422

Qy    318 acaagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttatt 377
          |||  | |  |  | || |||| |  |||||       ||  | || | |    |  |
Db    423 ACATCTTCTGATGAAATTTTCTAGAAGAATGGCAATTATGAACAATATGTCTTTGATCAT 482

Qy    378 tttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagag 437
             |    |||  |    |  |  |  |  |     |  |     || |  | |  |
Db    483 ACATAGGTCTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGGCACCTGCATGCTAC 542

Qy    438 ttcttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaata 497
          ||| |||||||  ||  |    || || ||| |   ||| |       |  | || || ||
Db    543 TTCCTTTGATGTGGTTCTAACAGACCCCTTTCACCTCTGCGCGGCGGTGCTGGCTAAGTA 602

Qy    498 tttctccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaagg 557
            | ||  | ||  | ||| | |||   |||   || |  ||  | |       | ||
Db    603 CCTGTCGATTCCTGCTGTGTTTTTCTTGAGGAACATTCCATGTGATTTAGACTTTAAGGG 662

Qy    558 tgcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgc 617
            |||||||  ||    |||   |||||| | ||  |||  | ||          |||||
Db    663 CACACAGTGTCCAAACCCTTCCTCCTATATTCCTAGATTACTAACGACCAATTCAGACCA 722

Qy    618 catgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatg 677
          ||||||  |||  | | ||  ||   ||||    ||   |   |||        |  | ||
Db    723 CATGACATTCCTGCAAAGGGTCAAGAACATGCTCTACCCTCTGGCCCTGTCCTACCTTTG 782

Qy    678 ccaccgttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttac 737
          |||   | |||       |     |    | ||||||||  || | || | |   ||  |
Db    783 CCATGCTGTTTCTGCTCCTTATGCAAGCCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTC 842

Qy    738 ggagtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttgga 797
            |  |   ||||  ||||||  ||||   ||| |||| |||  ||| |||||  ||||
Db    843 AGTGGTGGATCTTGTCAGCCATGCATCCTGGTGGCTGTTCCGAGGGGAGTTTGTGATGGA 902

Qy    798 ctatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggg 857
          || ||||  ||  | |||||||||||| |||||||||| || ||||||||     | ||
Db    903 TTACCCCAGGCCGATCATGCCCAACATGGTCTTCATTGGGGGCATCAACTGTGCCAACGG 962

Qy    858 aaagccgttgcctatggaatttgaagcctacatta 892
          |||||  |  ||  ||  | |   |    | |||
Db    963 GAAGCCACTATCTCAGGTCTGTATTGGTGCCTTTA 997


RESULT   7
PCT-US92-00282-16
; Sequence 16, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1448 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  YES
PCT-US92-00282-16


  Query Match             8.0%;  Score 186.6;  DB 5;  Length 1448;
  Best Local Similarity   53.8%;  Pred. No. 1.1e-43;
  Matches  429;  Conservative    0;  Mismatches  364;  Indels    4;  Gaps    2;

Qy     81 ctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcaccat 140
          ||  || |||   |||||| | ||||| || ||||| ||||| |||||||||  ||  |||
Db    178 CTGGGCTGAGAGTGGGAAGGTGCTGGTGGTACCCATTGATGGCAGCCACTGGCTCG-CAT 236

Qy    141 gaggtcggtggtggagaaactcattctcaggggcatgaggtggttgtagtcatgccaga 200
          | ||  ||  |||  | | |||  |  ||| || ||  ||| ||| ||  |||  |||||
Db    237 GCGGGAGGCCGTGCGGGAGCTCCATGCCAGAGGCCACCAGGCGGTGGTCCTCACCCCAGA 296

Qy    201 ggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttcata 260
          ||||| |  |||  |   |  |        || |||  |||  || ||| | | ||||||
Db    297 GGTGAATATGCACATCAAACAAGAGAAATTTTTCACCCTGACAACCTATGCCATTTCATG 356

Qy    261 taccctggaggatctggaccgggagttcaaggcttttgcccatgctcaat---ggaaagc 317
          ||||| ||| ||  | || ||     |   ||       | ||     | |       | ||
Db    357 GACCCAGGATGAATTTGATCGCCTTGTGCTGGGCCACACTCAACTGTACTTTGAAACAGA 416

Qy    318 acaagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttatt 377
          |||  | |  |  | || ||||  |   ||||      |  ||  |    | |    | ||
Db    417 ACATTTTCTGAAGACATTTTCTAAAAGTATGGCAATTTTGAAAAATTCATCTTTGGTCTT 476

Qy    378 tttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagag 437
          |  |     |||   |  |  |  |  |     |  |     || | || |  |
Db    477 TCATAGGTCTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGGCACCTGAATGCTAC 536

Qy    438 ttcttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaata 497
          ||| || ||||  || ||  | || ||  || |   ||||||   | |  | || || ||
Db    537 TTCCTTCGATGTGGTTTTAATAGACCCAATTTACCTCTGTGGGGCAGTGCTGGCTAAGTA 596

Qy    498 tttctccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaagg 557
            | ||  | ||  | ||| | ||    |||   || |  ||  | |       | ||
Db    597 CCTATCAATTCCTGCTGTGTTTTTTTTGAGGAACATTCCATGTGATTTAGACTTTAAGGG 656

Qy    558 tgcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgc 617
            |||||||  ||   ||||   |||||| | || |   | ||          |||||
Db    657 CACACAGTGTCCAAATCCTTCCTCCTATATTCCTAAGTTACTAACGACCAATTCAGACCA 716

Qy    618 catgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatg 677
          ||||||  |||  | | ||  ||   ||||    ||   |   |||        |  | ||
Db    717 CATGACATTCCTGCAAAGGGTCAAGAACATGCTCTACCCTCTGGCCCTGTCCTACCTTTG 776

Qy    678 ccaccgttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttac 737
          |||   | |||       |     |    | ||||||||  || | || | |   ||  |
Db    777 CCATGCTCTTTCTGCTCCTTATGCAAGCCTTGCCTCTGAGCTTTTTCAGAGGGAGGTGTC 836

Qy    738 ggagtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttgga 797
            | |   |||||  ||||||  ||||  | ||| |||| |||  |||||||||  ||||
Db    837 AGTGGTGGATCTTCTCAGCCATGCATCTGTGTGGCTGTTCCGAGGGGACTTTGTGATGGA 896

Qy    798 ctatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggg 857
            || ||||  ||  | |||||||||||| |||||||||| || ||||||||     | ||
Db    897 TTACCCCAGGCCGATCATGCCCAACATGGTCTTCATTGGGGGCATCAACTGTGCCAACAG 956
```

```
Qy   1251 tgttctggaaatgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaa 1310
          |   |   ||| |   |  ||||||    |||||  | ||  |||||||| ||||||
Db   1302 catcaggaccatgtcaagtagagatttgctcaatgcattgaagtcagtcattaatgaccc 1361

Qy   1311 aagttacaaggagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagcc 1370
          |  || || ||||| ||||||    | || ||  ||||  | || |  |||||| ||||
Db   1362 tatctataaagagaatatcatgaaattatcaagaattcatcatgatcaaccggtgaagcc 1421

Qy   1371 gctggacctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacct 1430
          ||||| |  || || |||||| | |||||||| ||| | || || || ||    |||||
Db   1422 cctggatcgagcagtcttctggattgagtttgtcatgcgccataaaggagccaagcacct 1481

Qy   1431 gcgccccgcagcccacgacctcacctggtaccagtaccattccttggacgtgattggttt 1490
          ||   |||||||||| ||||||||||  |||||||| || ||||| ||||| |  ||
Db   1482 tcgggtcgcagcccacaacctcacctggatccagtaccactctttggatgtgatagcatt 1541

Qy   1491 cctcttggccgtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggcta 1550
          ||| ||||| |||| || ||  || || | |  ||||||||    | | | |
Db   1542 cctgctggcctgcgtggcaactatgatatttatgatcacaaaatgttgcctgttttgttt 1601

Qy   1551 ccggaa 1556
          ||| ||
Db   1602 ccgaaa 1607


RESULT   5
PCT-US92-00282-22
; Sequence 22, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  783 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-22


  Query Match             9.6%;  Score 222.4;  DB 5;  Length 783;
  Best Local Similarity   99.6%;  Pred. No. 3.9e-54;
  Matches  223;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1093 ggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgtttatgaaagc 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    471 GGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAGC 530

Qy   1153 atatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatggacaatgca 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    531 ATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAATGCA 590

Qy   1213 aagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatgacttctgaa 1272
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    591 AAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGAA 650

Qy   1273 gatttagaaaaatgctctaaaagcagtcatcaatgacaaaagtta 1316
          |||||||| |||||||||||||||||||||||||||||||||||
Db    651 GATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGGTA 694


RESULT   6
PCT-US92-00282-10
; Sequence 10, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1008 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-10


  Query Match             8.3%;  Score 192.6;  DB 5;  Length 1008;
  Best Local Similarity   53.1%;  Pred. No. 1.7e-45;
  Matches  433;  Conservative    0;  Mismatches  379;  Indels    3;  Gaps    1;

Qy     81 ctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcaccat 140
          ||  || |||   |||||| | ||||| |||||||  ||||| ||||||||| ||| |||
Db    183 CTGGGCTGAGAGTGGGAAGGTGCTGGTGGTGCCCACTGATGGCAGCCACTGGCTCAGCAT 242

Qy    141 gaggtcggtggtggagaaactcattctcaggggcatgaggtggttgtagtcatgccaga 200
          | ||  ||   ||  | | |||  |   || || ||  ||||||| ||  |||  |  ||
Db    243 GCGGGAGGCCTTGCGGGACCTCCATGCGAGAGGCCACCAGGTGGTGGTCCTCACCCTGGA 302

Qy    201 ggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttcata 260
          ||||| |  | |  |   |  |        || |||  |||  ||  ||| | | |||||
```

```
Db    847 TTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCA 906

Qy   1949 ccgctgcccctactgcaaatggcagttttaatcttatcttttggcttctgcagatggttg 2008
Db    907 CGGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTTG 966

Qy   2009 caattgatcctaaccaataatggtcagtcctcatctctgtcctgcttcataggtgccac 2068
Db    967 CAATTGATCCTTAACCAATAATGGTCAGTCCTCATCTCTGTCGTGCTTCATAGGTGCCAC 1026

Qy   2069 cttgtgtgtttaaagaagggaagctttgtacctttagagtgtaggtgaaatgaatgaatg 2128
Db   1027 CTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATG 1086

Qy   2129 gcttggagtgcactgagaacagcatatgatttcttgctttggggaaaaagaatgatgcta 2188
Db   1087 GCTTGGAGTGCACTGAGAACAGCATATGATTTCTTGCTTTGGGGAAAAAGAATGATGCTA 1146

Qy   2189 tgaaattggtgggtggtgtatttgagaagataatcattgcttatgtcaaatggagctgaa 2248
Db   1147 TGAAATTGGTGGGTGGTGTATTTGAGAAGATAATCATTGCTTATGTCAAATGGAGCTGAA 1206

Qy   2249 tttgataaaaacccaaaatacagctatgaagtgctgggcaagtttacttttttttctgatg 2308
Db   1207 TTTGATAAAAACCCAAAATACAGCTATGAAGTGCTGGGCAAGTTTACTTTTTTTCTGATG 1266

Qy   2309 tttcctacaact 2320
Db   1267 TTTCCTACAACT 1278
```

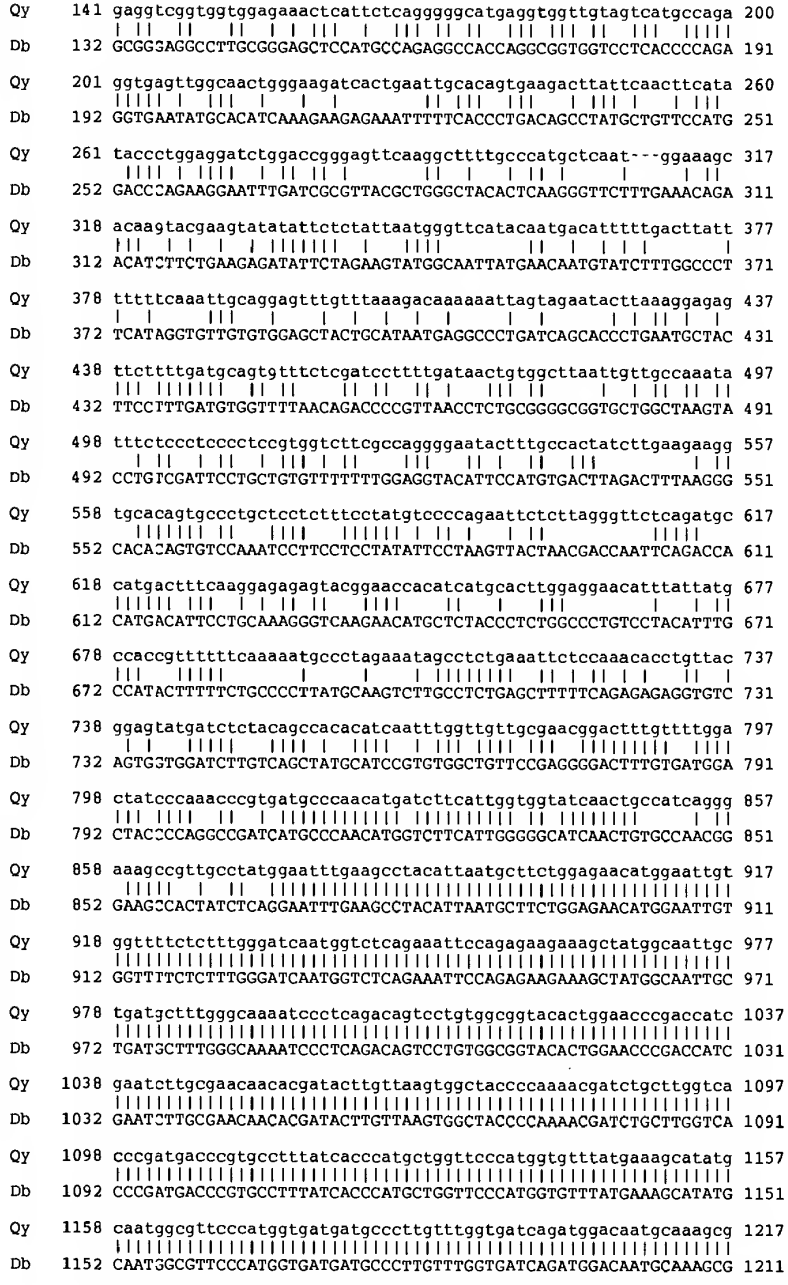
RESULT 4
US-09-180-852-1
; Sequence 1, Application US/09180852
; Patent No. 6287834
; GENERAL INFORMATION:
; APPLICANT: BELANGER, Alain
; APPLICANT: HUM, Dean W.
; APPLICANT: BEAULIEU, Martin
; APPLICANT: LEVESQUE, Eric
; TITLE OF INVENTION: CHARACTERIZATION AND USE OF AN ISOLATED URIDINE
; TITLE OF INVENTION: DIPHOSPHO-GLUCURONOSYLTRANSFERASE
; FILE REFERENCE: 1259-449
; CURRENT APPLICATION NUMBER: US/09/180,852
; CURRENT FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: PCT/CA97/00328
; EARLIER FILING DATE: 1997-05-16
; EARLIER APPLICATION NUMBER: US 08/649,319
; EARLIER FILING DATE: 1996-05-17
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2107
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52)..(1644)
US-09-180-852-1

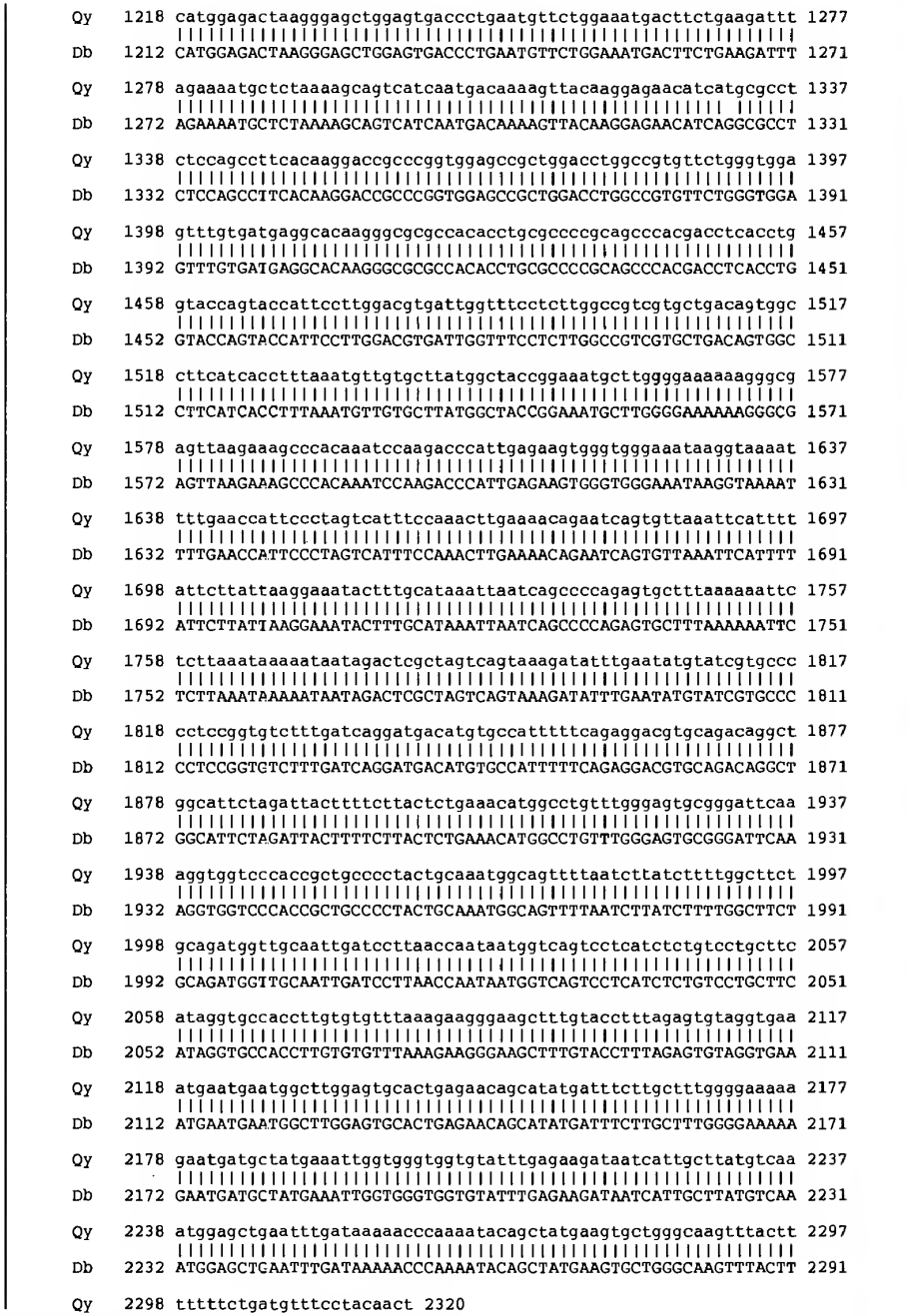
Query Match 14.8%; Score 344.4; DB 4; Length 2107;
Best Local Similarity 53.9%; Pred. No. 5.6e-89;
Matches 811; Conservative 0; Mismatches 671; Indels 24; Gaps 4;

```
Qy     75 ctgtggctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggtt 134
Db    102 ctttagctctgggagttgtggaaaggtgctggtgtggcccacagaatacagccattggat 161

Qy    135 caccatgaggtcggtggtggagaaactcattctcaggggcatgaggtggttgtagtcat 194
Db    162 aaatatgaagacaatcctggaagagcttgttcagaggggtcatgaggtgattgtgttgac 221

Qy    195 gccagaggtgagttggcaactgg------gaagatcactgaattgcacagtgaagactta 248
Db    222 atcttcggcttctattcttgtcaatgccagtaaatcatctgctattaaattagaagttta 281

Qy    249 ttcaacttcatataccctggaggatctggaccggga-----------gttcaaggcttt 296
Db    282 tcctacatctttaactaaaaatgatttggaagatttttttatgaaaatgttcgatagatg 341

Qy    297 tgcccatgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcata 356
Db    342 gacatatagtatttcaaaaaatacattttggtcatatttttcacaactacaagaattgtg 401

Qy    357 caatgacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaatt 416
Db    402 ttgggaatattctgactataatataaagctctgtgaagatgcagttttgaacaagaaact 461

Qy    417 agtagaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactg 476
Db    462 tatgagaaaactacaagagtcaaaatttgatgtccttctggcagatgccgttaatccctg 521

Qy    477 tggcttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatact 536
Db    522 tggtgagctgctggctgaactacttaacataccctttctgtacagtctccgcttctctgt 581

Qy    537 ttgccactatcttg---aagaaggtgcacagtgccctgctcctctttcctatgtccccag 593
Db    582 tggctacacagttgagaagaatggtggaggatttctgttccctccttcctatgtacctgt 641

Qy    594 aattctcttagggttctcagatgccatgactttcaaggagagagtacggaaccacat--- 650
Db    642 tgttatgtcagaattaagtgatcaaatgattttcatggagaggataaaaaatatgatata 701

Qy    651 catgcacttggaggaacatttattatgccaccgtttttttcaaaaatgccctagaaatagc 710
Db    702 tatgctttattttgacttttggtttcaagcatatgatctgaagaagtgggaccagtttta 761

Qy    711 ctctgaaattctccaaacacctgttacggagtatgatctctacagccacacatcaatttg 770
Db    762 tagtgaagttctaggaagacccactacattatttgagacaatggggaaagctgaaatgtg 821

Qy    771 gttgttgcgaacggacttttgttttggactatcccaaacccgtgatgcccaacatgatctt 830
Db    822 gctcattcgaacctattgggattttgaatttcctcgcccattcttaccaaatgttgattt 881

Qy    831 cattggtggtatcaactgccatcagggaaagccgttgcctatggaatttgaagcctacat 890
Db    882 tgttggaggacttcactgtaaaccagccaaacccttgcctaaggaaatggaagagtttgt 941

Qy    891 taatgcttctggagaacatggaattgtggttttctctttgggatcaatggtctcagaaat 950
Db    942 gcagagctctggagaaaatggtattgtggtgttttctctggggtcgatgatcagtaacat 1001

Qy    951 tccagagaagaaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcct 1010
Db   1002 gtcagaagaaagtgccaacatgattgcatcagcccttgcccagatcccacaaaaggttct 1061

Qy   1011 gtggcggtacactggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtg 1070
Db   1062 atggagatttgatggcaagaagccaaatactttaggttccaatactcgactgtataagtg 1121

Qy   1071 gctaccccaaaacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctgg 1130
Db   1122 gttaccccagaatgaccttcttggtcatcccaaaaccaaagcttttataactcatggtgg 1181

Qy   1131 ttcccatggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtt 1190
Db   1182 aaccaatggcatctatgaggcgatctaccatgggatccctatggtgggcattcccttgtt 1241

Qy   1191 tggtgatcagatggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaa 1250
Db   1242 tgcggatcaacatgataacattgctcacatgaaagccaagggagcagccctcagtgtgga 1301
```

```
Qy   1505 tgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttgg 1564
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1496 TGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGG 1555

Qy   1565 ggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtggga 1624
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1556 GGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGA 1615

Qy   1625 aataaggtaaaatttttgaaccattccctagtcatttccaaacttgaaaacagaatcagtg 1684
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1616 AATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTG 1675

Qy   1685 ttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtg 1744
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1676 TTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTG 1735

Qy   1745 ctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaat 1804
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1736 CTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAAT 1795

Qy   1805 atgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggac 1864
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1796 ATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGAC 1855

Qy   1865 gtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttggga 1924
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1856 GTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGA 1915

Qy   1925 gtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatctta 1984
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1916 GTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTA 1975

Qy   1985 tcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatc 2044
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1976 TCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCTCATC 2035

Qy   2045 tctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtaccttta 2104
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2036 TCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTTA 2095

Qy   2105 gagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttcttg 2164
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2096 GAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTTCTTG 2155

Qy   2165 ctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatca 2224
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2156 CTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATAATCA 2215

Qy   2225 ttgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgctg 2284
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2216 TTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGTGCTG 2275

Qy   2285 ggcaagtttacttttttttctgatgtttcctacaact 2320
          ||||||||||||||||||||||||||||||||||||
Db   2276 GGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 2311


RESULT   3
PCT-US92-00282-25
; Sequence 25, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1561 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-25


  Query Match             41.7%;  Score 967.2;  DB 5;  Length 1561;
  Best Local Similarity   98.0%;  Pred. No. 3.3e-268;
  Matches 1011;  Conservative    0;  Mismatches   18;  Indels    3;  Gaps    3;

Qy   1291 aaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctccagccttcac 1350
          || | |    ||  ||     ||||||||||||||||||||||||||||||||||||||||||
Db    248 AACGAAACTGTCTTTGTGTTTAGTTACAAGGAGAACATCATGCGCCTCTCCAGCCTTCAC 307

Qy   1351 aaggaccgcccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgagg 1410
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    308 AAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGG 367

Qy   1411 cacaagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccat 1470
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    368 CACAAGGGCGCGCCACACCTGCGCCCCGCAGCCCACGACCTCACCTGGTACCAGTACCAT 427

Qy   1471 tccttggacgtgattggtttcctcttggccgtcgtgct-gacagtggccttcatcacctt 1529
          ||||||||||||||||||||||||||||||||||||  ||| ||||||||||||||||||||
Db    428 TCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCTGGCTGGACAGTGGCCTTCATCACCTT 487

Qy   1530 taaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcgagttaagaaagc 1589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    488 TAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGGAAAAAAGGGCGAGTTAAGAAAGC 547

Qy   1590 ccacaaatccaagacccattgagaagtgggtgggaaataaggtaaaattttgaaccattc 1649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    548 CCACAAATCCAAGACCCATTGAGAAGTGGGTGGGAAATAAGGTAAAATTTTGAACCATTC 607

Qy   1650 cctagtcatttccaaacttgaaaacagaatcagtgttaaattcattttattcttattaag 1709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    608 CCTAGTCATTTCCAAACTTGAAAACAGAATCAGTGTTAAATTCATTTTATTCTTATTAAG 667

Qy   1710 gaaatactttgcataaattaatcagccccagagtgctttaaaaaattctcttaaataaaa 1769
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    668 GAAATACTTTGCATAAATTAATCAGCCCCAGAGTGCTTTAAAAAATTCTCTTAAATAAAA 727

Qy   1770 ataatagactcgctagtcagtaaagatatttgaatatgtatcgtg-ccccctccggtgtc 1828
          ||||||||||||||||||||||||||||||||||||||||||| ||||||| ||||||
Db    728 ATAATAGACTCGCTAGTCAGTAAAGATATTTGAATATGTATCGTGCCCCCCTCTGGTGTC 787

Qy   1829 tttgatcaggatgacatgtgccatttttcagaggacgtgcagacaggctggcattctaga 1888
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    788 TTTGATCAGGATGACATGTGCCATTTTTCAGAGGACGTGCAGACAGGCTGG-ATTCTAGA 846

Qy   1889 ttacttttcttactctgaaacatggcctgtttgggagtgcgggattcaaaggtggtccca 1948
```

```
          ||||||||||||||||||||||||
Db   2292 TTTTTCTGATGTTTCCTACAACT 2314


RESULT   2
PCT-US92-00282-1
; Sequence 1, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2336 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-1


  Query Match             69.4%;  Score 1609.2;  DB 5;  Length 2336;
  Best Local Similarity   82.7%;  Pred. No. 0;
  Matches 1865;  Conservative    0;  Mismatches  388;  Indels    3;  Gaps    2;

Qy      68 tgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatgggagcc 127
           |||||  |   |   |  || | || ||||| ||||| |  | ||  |||||||  ||||
Db      56 TGCTGGGCCCAGTGGTGTCCCATGCTGGGAAGATACTGTTGATCCCAGTGGATGGCAGCC 115

Qy     128 actggttcaccatgaggtcggtggtggagaaactcattctcagggggcatgaggtggttg 187
           |||||  | |  ||||      ||    |  || | ||    |  ||||| |||||  | ||||
Db     116 ACTGGCTGAGCATGCTTGGGGCCATCCAGCAGCTGCAGCAGAGGGGACATGAAATAGTTG 175

Qy     188 tagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaagactt 247
           |  |    || || |     | | |  |  ||     |     || |||  ||||||| |
Db     176 TCCTAGCACCTGACGCCTCGTTGTACATCAGAGACGGAGCATTTTACACCTTGAAGACGT 235

Qy     248 attcaacttcatataccctggaggatctggaccgggagtt--caaggcttttgcccatgc 305
           |  |     |||       ||||||| || |   |   ||    || ||   ||  |
Db     236 ACCCTGTGCCATTCCAAAGGGAGGATGTGAAAGAGTCTTTTGTTAGTCTCGGGCATAATG 295

Qy     306 tcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgacat 365
           |   ||  ||  |    | |        | |  || |   ||       | | ||  |
Db     296 TTTTTGAGAATGATTCTTTCCTGCAGCGTGTGATCAAAACATACAAGAAAATAAAAAAGG 355

Qy     366 ttttgact-tatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaat 424
             |   || |  |||| ||    ||       ||  |  |  ||||  |  |  | |  |

Db     356 ACTCTGCTATGCTTTTGTCTGGCTGTTCCCACTTACTGCACAACAAGGAGCTCATGGCCT 415

Qy     425 acttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggcttaa 484
           | |    || ||    |||||||   || |    || |||||   |   ||  ||   |
Db     416 CCCTGGCAGAAAGCAGCTTTGATGTCATGCTGACGGACCCTTTCCTTCCTTGCAGCCCCA 475

Qy     485 ttgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgccact 544
           | || ||| | ||  | || || ||| | ||  |||||      | | | |  |||  |
Db     476 TCGTGGCCCAGTACCTGTCTCTGCCCACTGTATTCTTCTTGCATGCACTGCCATGCAGCC 535

Qy     545 atcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttag 604
               || | | ||||||||    ||  | |||||  || |||||   |||||
Db     536 TGGAATTTGAGGCTACCCAGTGCCCCAACCCATTCTCCTACGTGCCCAGGCCTCTCTCCT 595

Qy     605 ggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggagg 664
           ||||||  ||||||  |||  | || | ||   ||||    ||||   |||
Db     596 CTCATTCAGATCACATGACCTTCCTGCAGCGGGTGAAGAACATGCTCATTGCCTTTTCAC 655

Qy     665 aacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaattctcc 724
           |  | ||  | ||| ||    |||              |    | ||||| ||| | || |
Db     656 AGAACTTTCTGTGCGACGTGGTTTATTCCCCGTATGCAACCCTTGCCTCAGAATTCCTTC 715

Qy     725 aaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacgg 784
           | | |   || || |  | || || |  |||    ||||  | ||| ||||  ||| |
Db     716 AGAGAGAGGTGACTGTCCAGGACCTATTGAGCTCTGCATCTGTCTGGCTGTTTAGAAGTG 775

Qy     785 actttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatca 844
           |||||||   ||| || || |  ||| | |||||||| ||| | ||  |||||||  ||||
Db     776 ACTTTGTGAAGGATTACCCTAGGCCCATCATGCCCAATATGGTTTTTGTTGGTGGAATCA 835

Qy     845 actgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctggag 904
           |||||| |||   ||| ||  |  |  ||||||||||||||||||||||||||||||||||
Db     836 ACTGCCTTCACCAAAATCCACTATCCCAGGAATTTGAAGCCTACATTAATGCTTCTGGAG 895

Qy     905 aacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaagaaag 964
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     896 AACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAG 955

Qy     965 ctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactg 1024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     956 CTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTG 1015

Qy    1025 gaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacg 1084
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1016 GAACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACG 1075

Qy    1085 atctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgttt 1144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1076 ATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTT 1135

Qy    1145 atgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatgg 1204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1136 ATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGG 1195

Qy    1205 acaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatga 1264
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1196 ACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGA 1255

Qy    1265 cttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggaga 1324
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1256 CTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGA 1315

Qy    1325 acatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccg 1384
           ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1316 ACATCAGGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCG 1375

Qy    1385 tgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagccc 1444
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1376 TGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCC 1435

Qy    1445 acgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcg 1504
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1436 ACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCG 1495
```





GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 21, 2002, 02:52:32 ; Search time 148.28 Seconds
(without alignments)
3843.200 Million cell updates/sec

Title: US-09-835-082-1
Perfect score: 2320
Sequence: 1 ctcagctgcagttctctgat.....ttctgatgtttcctacaact 2320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1633.4	70.4	2339	5	PCT-US92-00282-2	Sequence 2, Appli
	2	1609.2	69.4	2336	5	PCT-US92-00282-1	Sequence 1, Appli
	3	967.2	41.7	1561	5	PCT-US92-00282-25	Sequence 25, Appl
	4	344.4	14.8	2107	4	US-09-180-852-1	Sequence 1, Appli
	5	222.4	9.6	783	5	PCT-US92-00282-22	Sequence 22, Appl
	6	192.6	8.3	1008	5	PCT-US92-00282-10	Sequence 10, Appl
	7	186.6	8.0	1448	5	PCT-US92-00282-16	Sequence 16, Appl
	8	183.6	7.9	1066	5	PCT-US92-00282-14	Sequence 14, Appl
	9	161.8	7.0	1190	5	PCT-US92-00282-18	Sequence 18, Appl
	10	161.2	6.9	1197	5	PCT-US92-00282-12	Sequence 12, Appl
	11	143.6	6.2	1219	5	PCT-US92-00282-8	Sequence 8, Appli
	12	135	5.8	513	5	PCT-US92-00282-20	Sequence 20, Appl
	13	54.6	2.4	1669	2	US-08-522-421-1	Sequence 1, Appli
	14	44.8	1.9	1607	2	US-08-797-226-1	Sequence 1, Appli
	15	43.6	1.9	51	5	PCT-US92-00282-37	Sequence 37, Appl
	16	39.8	1.7	7218	1	US-08-232-463-14	Sequence 14, Appl
	17	38.2	1.6	1627	3	US-09-106-464-1	Sequence 1, Appli
	18	36.6	1.6	1731	2	US-08-466-583-1	Sequence 1, Appli
	19	36.6	1.6	1731	5	PCT-US95-07820-1	Sequence 1, Appli
c	20	36.6	1.6	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	21	36.6	1.6	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	22	36	1.6	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	23	36	1.6	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	24	35.8	1.5	1193	4	US-09-372-422A-23	Sequence 23, Appl
c	25	35.8	1.5	1696	1	US-07-612-674-1	Sequence 1, Appli
c	26	35.2	1.5	792	3	US-08-950-720A-1	Sequence 1, Appli
	27	35.2	1.5	1636	6	5447867-2	Patent No. 5447867
c	28	35.2	1.5	4839	4	US-09-353-332-1	Sequence 1, Appli
c	29	35.2	1.5	10607	1	US-08-078-090-3	Sequence 3, Appli
	30	35	1.5	1800	6	5180581-1	Patent No. 5180581
	31	35	1.5	2793	1	US-08-281-916-5	Sequence 5, Appli
	32	35	1.5	2793	2	US-08-460-725-7	Sequence 7, Appli
	33	35	1.5	10815	4	US-09-004-838-21	Sequence 21, Appl
	34	35	1.5	13149	4	US-09-004-838-87	Sequence 87, Appl
	35	34.6	1.5	1330	2	US-08-933-750C-80	Sequence 80, Appl
	36	34.6	1.5	1330	3	US-09-234-613-80	Sequence 80, Appl
c	37	34.6	1.5	1534	1	US-08-480-784-13	Sequence 13, Appl
c	38	34.6	1.5	1534	1	US-08-483-553-13	Sequence 13, Appl
c	39	34.6	1.5	1534	1	US-08-487-002-13	Sequence 13, Appl
c	40	34.6	1.5	1534	1	US-08-483-554B-13	Sequence 13, Appl
c	41	34.6	1.5	1534	1	US-08-488-011B-13	Sequence 13, Appl
c	42	34.6	1.5	1534	4	US-08-850-727-13	Sequence 13, Appl
c	43	34.6	1.5	1534	5	PCT-US95-10202-13	Sequence 13, Appl
c	44	34.6	1.5	1534	5	PCT-US95-10203-13	Sequence 13, Appl
c	45	34.6	1.5	1534	5	PCT-US95-10220-13	Sequence 13, Appl

ALIGNMENTS

```
RESULT   1
PCT-US92-00282-2
; Sequence 2, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2339 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-2


  Query Match            70.4%;  Score 1633.4;  DB 5;  Length 2339;
  Best Local Similarity  83.3%;  Pred. No. 0;
  Matches 1869;  Conservative    0;  Mismatches  371;  Indels    3;  Gaps    1;

Qy      81 ctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcaccat 140
           ||  || |||   || ||| |  |||| |||||||  ||||| |||| |||| ||| |||
Db      72 CTGGGCTGAGAGTGGAAAGGTGTTGGTGGTGCCCACTGATGGCAGCCCCTGGCTCAGCAT 131
```

```
Qy    544 tatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctctta 603
          ||||||||||||||||||||||||||||||||||||||||||||||||||    |||||||
Db    558 CATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAATGATCTCTTA 617

Qy    604 gggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggag 663
          ||||||||||||||||||||||||||||||||||| |||||||||| ||||||||||
Db    618 GGGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTATGGAACCACATCGTGCACTTGGAG 677

Qy    664 gaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaattctc 723
          || |||||||| |||||    | || | | ||||||||||||||||| ||||||||||||
Db    678 GACCATTTATT-TGCCAGTATCTTATTAGAAATGCCCTAGAAATAG-CTCTGAAATTCTC 735

Qy    724 caaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacg 783
          ||||| ||||| |||| |||||||||||||| ||| ||||||||||||||||||||||
Db    736 CAAACCCCTGTCACGGCATATGATCTCTACAGTCACCCATCAATTTGGTTGTTGCGAACG 795

Qy    784 gactttgttttggac-tatcccaaacccgtgatg 816
          | | |||||| |||| ||| || |||||||| ||
Db    796 GGCCTTGTTTGGGACTTATTCCGAACCCGTGCTG 829


RESULT  15
BG400591
LOCUS       BG400591                 833 bp    mRNA    linear   EST 12-MAR-2001
DEFINITION  602464244F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4592315 5',
            mRNA sequence.
ACCESSION   BG400591
VERSION     BG400591.1  GI:13294039
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 833)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1330  row: h  column: 12
            High quality sequence stop: 800.
FEATURES             Location/Qualifiers
     source          1. .833
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4592315"
                     /clone_lib="NIH_MGC_75"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.65
                     kb (range 0.5-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      206 a    179 c    211 g    237 t
ORIGIN


  Query Match             27.6%;  Score 640;  DB 10;  Length 833;
  Best Local Similarity   91.1%;  Pred. No. 1.3e-156;
  Matches  760;  Conservative    0;  Mismatches   60;  Indels   14;  Gaps    7;
```

```
Qy      8 gcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtgtgtctgc 67
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGAGTTCTCTGATGGCTTGCACAGGGTGGACCAGCCCCCTTCCTCTATGTGTGTGTCTGC 60

Qy     68 tg--ctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatgggag 125
          ||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGTCGTGACCTGTGGCTTTGCCGAGGCAGGGAAGCTACTGGTAGTGCCCATGGATGGGAG 120

Qy    126 ccactggttcaccatgaggtcggtggtggagaaactcattctcaggggggcatgaggtggt 185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCACTGGTTCACCATGAGGTCGGTGGTGGAGAAACTCATTCTCAGGGGGCATGAGGTGGT 180

Qy    186 tgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaagac 245
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGTAGTCATGCCAGAGGTGAGTTGGCAACTGGGAAGATCACTGAATTGCACAGTGAAGAC 240

Qy    246 ttattcaacttcatatacccctggaggatctggaccgggagttcaaggcttttgcccatgc 305
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTATTCAACTTCATATACCCTGGAGGATCTGGACCGGGAGTTCAAGGCTTTTGCCCATGC 300

Qy    306 tcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgacat 365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCAATGGAAAGCACAAGTACGAAGTATATATTCTCTATTAATGGGTTCATACAATGAGCA 360

Qy    366 ttttg----acttatttttttcaaattgcaggagtttgtttaaagacaaa-aaattagta 420
          |||||       | ||| ||  || ||||||||||||||||||||||||| |||||||||
Db    361 TTTTGTGACTTATTTTTGTTCAAAGTTGCAGGAGTTTGTTTAAAGACAAAGAAATTAGTA 420

Qy    421 gaatact-taaaggagagttcttttgatgcagtg---tttctcgatccttttgataactg 476
          ||||||| |||||||||||||| ||||||||||   | ||  |  | |||    |  |
Db    421 GAATACTGTAAAGGAGAGTTCTGTTGATGCAGTGTGATCTCGAGTCCGTTTGATAACTGG 480

Qy    477 tggcttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatact 536
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGGCTTAATGGTTGCCAAATATTTCTCCCTCCCCTCCGTGGTCTTCGCCAGGGGAATACT 540

Qy    537 ttgccactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaat 596
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTGCCACTATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAAT 600

Qy    597 tctcttagcgttctcagatgccatgactttcaaggagagagtacggaaccacatcatgca 656
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TCTCTTAGCGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTACGGAACCACATCATGCA 660

Qy    657 cttgg--acgaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctct 714
          |||||  |||||||||| |||||||||||| || |  |||     |    ||    ||||
Db    661 CTTGGGAACGAACATTGATTATGCCACCGTGTTGTTCAAAGAATGCCTAGAAATAGCTCT 720

Qy    715 gaaattctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttg 774
          ||||||||||||||||||||||||||||||||||    |     |  | |||||||
Db    721 GAAATTCTCCAAACACCTGTTACGGAGTATGATCTCTAACAGCCACACTTCAATTGGTTG 780

Qy    775 ttgcgaacggactttgttttggactatcccaaacccgtgatgcccaacatgatc 828
          ||||||||||||||| |||||||||||||||||| ||||||||||||||||
Db    781 TTGCGAACCGACTTTG-TTTGGACTATCCCAAACCCGGAATGCCCAACATGATC 833


Search completed: August 21, 2002, 09:38:48
Job time: 24936 sec
```

```
ORIGIN


  Query Match             27.8%;  Score 644;  DB 10;  Length 723;
  Best Local Similarity   96.7%;  Pred. No. 1.1e-157;
  Matches  700;  Conservative    0;  Mismatches   20;  Indels    4;  Gaps    4;

Qy   1078 caaaacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccat 1137
          ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db      2 CAAAACGATCTGCTTGGTCACCCGATGACCCGTG-CTTTATCACCCATGCTGGTTCCCAT 60

Qy   1138 ggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgat 1197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGTGTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGAT 120

Qy   1198 cagatggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctg 1257
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CAGATGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTG 180

Qy   1258 gaaatgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttac 1317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GAAATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTAC 240

Qy   1318 aaggagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggac 1377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAGGAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGAC 300

Qy   1378 ctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgcccc 1437
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCC 360

Qy   1438 gcagcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttg 1497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCAGCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTG 420

Qy   1498 gccgtcgtgctgacagtggccttcatcaccttaaatgttgtgcttatggctaccggaaa 1557
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GCCGTCTTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAA 480

Qy   1558 tgcttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtg 1617
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGCTTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTG 540

Qy   1618 ggtgggaaataaggtaaaa-ttttgaaccattccctagtcatttccaaacttgaaaacag 1676
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    541 GGTGGGAAATAAGGTAAAATTTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAG 600

Qy   1677 aatcagtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagcc 1736
          |||||||||| |||||||||||||||||||||||| |||||||||||||||| |
Db    601 AATCAGTGTTACATTCATTTTATTCTTATTAAGGAAATAC-TTGCATAAATTAATCAG-C 658

Qy   1737 ccagagtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagat 1796
          |||||||||||||||||||||||||  ||| |||      | |  ||| ||    |||
Db    659 CCAGAGTGCTTTAAAAAATTCTCTTAATAAAATATAGACTCGTAGTCAGTAAGATTTGAT 718

Qy   1797 attt 1800
          || |
Db    719 ATGT 722


RESULT  14
BG469506
LOCUS       BG469506                 829 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602532818F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4660540 5',
            mRNA sequence.
ACCESSION   BG469506
VERSION     BG469506.1  GI:13401781
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 829)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1458  row: c  column: 05
            High quality sequence stop: 795.
FEATURES             Location/Qualifiers
     source          1. .829
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4660540"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      186 a    190 c    207 g    246 t
ORIGIN


  Query Match             27.6%;  Score 641.2;  DB 10;  Length 829;
  Best Local Similarity   89.9%;  Pred. No. 6.4e-157;
  Matches  732;  Conservative    0;  Mismatches   78;  Indels    4;  Gaps    4;

Qy      5 gctgcagttctctgatggcttgcacagggtggaccagccccctcctctatgtgtgtgtc 64
          |||||||||||||| |||||| || |||||||||||||||||| ||||| ||||||||||||
Db     18 GCTGCAGTTCTCTCATGGCTCGCGCAGGGTGGACCAGCCCCGTTCCTTTATGTGTGTC 77

Qy     65 tgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatggga 124
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     78 TACTGCTGACCTGTGGCTTTGCCGAGGCAGGGAAGCTGCTGGTAGTGCCCATGGATGGGA 137

Qy    125 gccactggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgaggtgg 184
          | ||||||||||||||  |||||||||||||||||| || ||||||||||||||||||
Db    138 GTCACTGGTTCACCATGCAGTCGGTGGTGGAGAAACTTATCCTCAGGGGGCATGAGGTGG 197

Qy    185 ttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaaga 244
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    198 TTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGAAAGATCACTGAATTGCACAGTGAAGA 257

Qy    245 cttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcccatg 304
          |||| ||||| || || || ||||| |||| | ||||||| |||| || ||| |||||||
Db    258 CTTACTCAACCTCGTACACTCTGGAAGATCAGAACCGGGAATTCATGGTTTTCGCCCATG 317

Qy    305 ctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgaca 364
          |||||||||||||||| | || |||||||| |||||||||||| |||||| || ||
Db    318 CTCAATGGAAAGCACAGGCACAAAGTATATTTTCTCTATTAATGAGTTCATCCAGTGGTT 377

Qy    365 tttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaat 424
          || ||||||||||||||  |||||||||||| |||||| |||  ||||||||||||||
Db    378 TTCTTGACTTATTTTTTTCGCATTGCAGGAGTTGGTTTAATGACCGAAAATTAGTAGAAT 437

Qy    425 acttaaaggagagttcttttgatgcagtgtttc-tcgatccttttgataactgtggctta 483
          |||||||||||||||||| |||||||||||| | |||||||||||| ||||||||||
Db    438 ACTTAAAGGAGAGTTCTTTGGATGCAGTGTTTCGTGGATCCTTTTGATACCTGTGGCTTA 497

Qy    484 attgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgccac 543
          ||||| || |||||||||||||||||| |||||||| |||||||||| || |||||
Db    498 ATTGTCGCTAAATATTTCTCCCTCCCCTCTGTGGTCTTCACCAGGGGAATATTTGGCCAC 557
```

```
ACCESSION   AI831562
VERSION     AI831562.1  GI:5452233
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 764)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1491   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 454.
FEATURES             Location/Qualifiers
     source          1. .764
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2405161"
                     /clone_lib="NCI_CGAP_Lu19"
                     /tissue_type="squamous cell carcinoma, poorly
                     differentiated (4 pooled tumors, including primary and
                     metastatic)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     pooled lung tumor tissue, and was then primed with a Not I
                     - oligo(dT) primer. Double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I  and Eco RI sites of the modified
                     pT7T3 vector.  Library went througb one round of
                     normalization. Library constructed by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT      242 a    156 c    133 g    232 t      1 others
ORIGIN
```

```
  Query Match           28.9%;  Score 669.4;  DB 9;  Length 764;
  Best Local Similarity   96.8%;  Pred. No. 2.5e-164;
  Matches  693;  Conservative    0;  Mismatches   22;  Indels    1;  Gaps    1;

Qy   1605 ccattgagaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttccaa 1664
           ||  |||  |||||||||||||||| |||||| |||||||||| ||   ||||||||||
Db    755 CCCATGAAGAGTGGGTGGGAAATAGGGTAAATTTTTGAACCATCCCTACGTCATTTCCAA 696

Qy   1665 acttgaaaacagaatcagtgttaaattcattttattcttattaaggaaatactttgcata 1724
           |||||||    |   |||||||||||||| |||||||||||||| ||||||||||||||
Db    695 CCTTGAAACAGCATCAAGTGTTAAATTCATCTTATTCTTATTAAGAAAATACTTTGCATA 636

Qy   1725 aattaatcagccccagagtgctttaaaaaattctcttaaataaaaataatagactcgcta 1784
           |||||||||||||||||||| ||||| ||||||||||||||||||||||||||||||||
Db    635 AATTAATCAGCCCCAGAGTGC-TTAAANAATTCTCTTAAATAAAAATAATAGACTCGCTA 577

Qy   1785 gtcagtaaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgaca 1844
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    576 GTCAGTACAGATATTTGAATATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACA 517

Qy   1845 tgtgccatttttcagaggacgtgcagacaggctggcattctagattacttttcttactct 1904
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    516 TGTGCCATTTTTCAGAGGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCT 457
```

```
Qy   1905 gaaacatggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctactgc 1964
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    456 GAAACATGGCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGC 397

Qy   1965 aaatggcagttttaatcttatcttttggcttctgcagatggttgcaattgatccttaacc 2024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    396 AAATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACC 337

Qy   2025 aataatggtcagtcctcatctctgtcctgcttcataggtgccaccttgtgtgtttaaaga 2084
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    336 AATAATGGTCAGTCCTCATCTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGA 277

Qy   2085 agggaagctttgtacctttagagtgtaggtgaaatgaatgaatggcttggagtgcactga 2144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    276 AGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGA 217

Qy   2145 gaacagcatatgatttcttgctttggggaaaaagaatgatgctatgaaattggtgggtgg 2204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    216 GAACAGCATATGATTTCTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGG 157

Qy   2205 tgtatttgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaacccaa 2264
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    156 TGTATTTGAGAAGATAATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAA 97

Qy   2265 aatacagcatatgaagtgctgggcaagtttacttttttttctgatgtttcctacaact 2320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     96 AATACAGCTATGAAGTGCTGGGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 41
```

```
RESULT  13
BG424452
LOCUS       BG424452                723 bp    mRNA    linear   EST 14-MAR-2001
DEFINITION  602447772F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4586368 5',
            mRNA sequence.
ACCESSION   BG424452
VERSION     BG424452.1  GI:13330958
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 723)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1314  row: p  column: 17
            High quality sequence stop: 711.
FEATURES             Location/Qualifiers
     source          1. .723
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4586368"
                     /clone_lib="NIH_MGC_14"
                     /tissue_type="renal cell adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      202 a    160 c    170 g    191 t
```

```
Db    181 GGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGAAGATTTAGA 240

Qy   1281 aaatgctctaaaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctc 1340
Db    241 AAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAACATCATGCGCCTCTC 300

Qy   1341 cagccttcacaaggaccgcccggtggagccgctggacctggccgtgttctgggtggagtt 1400
Db    301 CAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAGTT 360

Qy   1401 tgtgatgaggcacaagggcgcgccacacctgcgccccgcagcccacgacctcacctggta 1460
Db    361 TGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCCACGACCTCACCTGGTA 420

Qy   1461 ccagtaccattccttggacgtgattggtttcctcttggccgtcgtgctgacagtggcctt 1520
Db    421 CCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGCCTT 480

Qy   1521 catcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcgagt 1580
Db    481 CATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGGAAAAAAGGGCGAGT 540

Qy   1581 taagaaagcccacaaatccaagacccattgagaagtgggtgggaaataaggtaaaatttt 1640
Db    541 TAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGAAATAAGGTAAAATTTG 600

Qy   1641 gaaccattccct-agtcatttccaaacttgaaaacagaatcagtgttaaattcattttat 1699
Db    601 GAACCATTCCCTAAGTCATTTCCAAACTTGAAAACAGAATCAGTGTTCAATTCATCT--A 658

Qy   1700 tcttattaaggaaatactttgcataaattaatcagccccagagtgctttaaaaaattctc 1759
Db    659 TCTTATTACGGAAATACTTCGCCTACATT-ATCAGGCCCAGAGTGCTTT-ACAAATCCTC 716

Qy   1760 ttaaataaaaataatagactcgctagtcagtaaagatattt 1800
Db    717 TAAATAACAAATAATAGACTCGCTAGTCAGAACAGATCTCT 757


RESULT  11
BG747143
LOCUS       BG747143                697 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602704364F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4857831 5',
            mRNA sequence.
ACCESSION   BG747143
VERSION     BG747143.1  GI:14057796
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1710  row: o  column: 16
            High quality sequence stop: 684.
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4857831"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      177 a    168 c    184 g    168 t
ORIGIN


  Query Match             29.3%;  Score 679.2;  DB 10;  Length 697;
  Best Local Similarity   99.4%;  Pred. No. 6.5e-167;
  Matches  692;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy    910 ggaattgtggttttctctttgggatcaatggtctcagaaattccagagaagaaagctatg 969
Db      2 GGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATG 61

Qy    970 gcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactggaacc 1029
Db     62 GCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGGAACC 121

Qy   1030 cgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacgatctg 1089
Db    122 CGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACGATCTG 181

Qy   1090 cttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgtttatgaa 1149
Db    182 CTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAA 241

Qy   1150 agcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatggacaat 1209
Db    242 AGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAAT 301

Qy   1210 gcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatgacttct 1269
Db    302 GCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCT 361

Qy   1270 gaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggagaacatc 1329
Db    362 GAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAACATC 421

Qy   1330 atgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccgtgttc 1389
Db    422 ATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTC 481

Qy   1390 tgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagcccacgac 1449
Db    482 TGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGGCCACGAG 541

Qy   1450 ctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcgtgctg 1509
Db    542 CTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGTGCTG 601

Qy   1510 acagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaa 1569
Db    602 ACAGTGGGCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGGAAA 661

Qy   1570 aaagggcgagttaagaaagcccacaaatccaagacc 1605
Db    662 AAAGGGCGAGTTAAGAAAG-CCACAAATCCAAGACC 696


RESULT  12
AI831562/c
LOCUS       AI831562                764 bp    mRNA    linear   EST 21-DEC-1999
DEFINITION  wj39b01.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2405161 3'
            similar to gb:M57899 UDP-GLUCURONOSYLTRANSFERASE 1A PRECURSOR,
            MICROSOMAL (HUMAN);, mRNA sequence.
```

```
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM10401  row: n  column: 18
           High quality sequence stop: 763.
FEATURES             Location/Qualifiers
     source          1. .881
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4514129"
                     /clone_lib="NIH_MGC_93"
                     /tissue_type="transitional cell papilloma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      263 a    154 c    198 g    266 t
ORIGIN


  Query Match             30.0%;  Score 695.4;  DB 10;  Length 881;
  Best Local Similarity   97.7%;  Pred. No. 4e-171;
  Matches  769;  Conservative    0;  Mismatches   11;  Indels    7;  Gaps    6;

Qy   1537 tgtgcttatggctaccggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaa 1596
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGTGCTTATGGCTACCGGAAATGCTTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAA 60

Qy   1597 tccaagacccattgagaagtgggtgggaaataaggtaaaattttgaaccattccctagtc 1656
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCCAAGACCCATTGAGAAGTGGGTGGGAAATAAGGTAAAATTTTGAACCATTCCCTAGTC 120

Qy   1657 atttccaaacttgaaaacagaatcagtgttaaattcattttattcttattaaggaaatac 1716
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTTCCAAACTTGAAAACAGAATCAGTGTTAAATTCATTTTATTCTTATTAAGGAAATAC 180

Qy   1717 tttgcataaattaatcagccccagagtgctttaaaaaattctcttaaataaaaataatag 1776
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTTGCATAAATTAATCAGCCCCAGAGTGCTTTAAAAAATTCTCTTAAATAAAAATAATAG 240

Qy   1777 actcgctagtcagtaaagatatttgaatatgtatcgtgccccctccggtgtctttgatca 1836
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACTCGCTAGTCAGTAAAGATATTTGAATATGTATCGTGCCCCCTCCGGTGTCTTTGATCA 300

Qy   1837 ggatgacatgtgccatttttcagaggacgtgcagacaggctggcattctagattacttttt 1896
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGATGACATGTGCCATTTTTCAGAGGACGTGCAGACAGGCTGGCATTCTAGATTACTTTT 360

Qy   1897 cttactctgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcc 1956
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTTACTCTGAAACATGGCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCACCGCTGCC 420

Qy   1957 cctactgcaaatggcagttttaatcttatcttttggcttctgcagatggttgcaattgat 2016
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCTACTGCAAATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGAT 480

Qy   2017 ccttaaccaataatggtcagtcctcatctctgtcctgcttcataggtgccaccttgtgtg 2076
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CCTTAACCAATAATGGTCAGTCCTCATCTCTGTCCTGCTTCATAGGTGCCACCTTGTGTG 540

Qy   2077 tttaaagaagggaagctttgtacctttagagtgtaggtgaaatgaatgaatggcttggag 2136
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTTAAAGAAGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAG 600

Qy   2137 -tgcactgagaaca-gcatatgat-ttcttgctttggggaaaaagaatgatgctatgaaa 2193
            ||||||||||||| ||||||||| ||||||  ||||||||||||||||||||||||||
Db    601 TTGCACTGAGAACAGGCATATGATCTTCTTGGCTTGGGGAAAAAGAATGATGCTATGAAA 660

Qy   2194 ttggtgggtggtgtatttgagaagataatcattgcttatgtcaaatggagctgaatttga 2253
           |||||||||||||||| |||||||||||||||| ||||||| ||||||||||||||||||
Db    661 -TGGTGGGTGGTGTATTCGAGAAGATAATCATTGTTTATGTCCAATGGAGCTGAATTTGA 719

Qy   2254 taaaaacccaaaatacagctatgaagtgctgggcaagtttacttttttttctgatgtttcc 2313
           | |||||||||||||||| |||||||||||||  |  ||   ||||||||||||||||||
Db    720 T-AAAACCCAAAATACAGGTATGAAGTGCTGGG--AAGTTTACTTTTTTCTGATGTTTCC 776

Qy   2314 tacaact 2320
           |||||||
Db    777 TACAACT 783


RESULT  10
BE870911
LOCUS       BE870911                1005 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601448862F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3852831 5',
            mRNA sequence.
ACCESSION   BE870911
VERSION     BE870911.1  GI:10319687
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1005)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM9575  row: l  column: 16
           High quality sequence stop: 645.
FEATURES             Location/Qualifiers
     source          1. .1005
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3852831"
                     /clone_lib="NIH_MGC_65"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.8 kb.   Library constructed by Life
                     Technologies.  "
BASE COUNT      282 a    268 c    246 g    209 t
ORIGIN


  Query Match             29.3%;  Score 680.6;  DB 10;  Length 1005;
  Best Local Similarity   96.8%;  Pred. No. 3.2e-167;
  Matches  737;  Conservative    0;  Mismatches   19;  Indels    5;  Gaps    4;

Qy   1041 tcttgcgaacaacacgatacttgttaagtggctaccccaaaacgatctgcttggtcaccc 1100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACGATCTGCTTGGTCACCC 60

Qy   1101 gatgacccgtgcctttatcacccatgctggttcccatggtgtttatgaaagcatatgcaa 1160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAGCATATGCAA 120

Qy   1161 tggcgttcccatggtgatgatgcccttgtttggtgatcagatggacaatgcaaagcgcat 1220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAATGCAAAGCGCAT 180

Qy   1221 ggagactaagggagctggagtgaccctgaatgttctggaaatgacttctgaagatttaga 1280
```

```
Qy    683 gttttttcaaaaatgccctag--aaatagcctctgaaattctcc-aaacacctgttacgg 739
          ||||||||||||||||||||  ||    |||||||||| |||| ||||||||||||||| 
Db    660 GTTTTTTCAAAAATGCCCTAGGAAATAGCCCTCTGAAATNCTCCAAAACACCTGTTACGG 719

Qy    740 agtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttggact 799
          |||||||||||||||||||||||||| | || |  |    |  |   |||| |||||||| 
Db    720 AGTATGATCTCTACAGCCACACATCAACTCGGCTCGTTGCGAACGGACTTGTCTTGGACT 779

Qy    800 atcccaaacccgtgatgcccaacatgatcttcattggtggtatcaac--tgccatcaggg 857
          || |||||||||||||||||  |||||| || |||| |||||| ||    |    | || 
Db    780 AT-CCAAACCCGTGATGCCCACAATGATCCTCCTTGGGGGTATCCACCTGCCGTCTACGG 838

Qy    858 aaagccgttgcctatggaatttga 881
          ||||||||||||||||| || || 
Db    839 AAAGCCGTTGCCTATGGATTTGGA 862
```

```
RESULT   8
AI307269/c
LOCUS       AI307269                818 bp    mRNA    linear   EST 08-APR-1999
DEFINITION  tb24d12.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2055287 3'
            similar to gb:M57899 UDP-GLUCURONOSYLTRANSFERASE 1A PRECURSOR,
            MICROSOMAL (HUMAN);, mRNA sequence.
ACCESSION   AI307269
VERSION     AI307269.1  GI:4002458
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 818)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 453.
FEATURES             Location/Qualifiers
     source          1. .818
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2055287"
                     /clone_lib="NCI_CGAP_Kid12"
                     /tissue_type="2 pooled tumors (clear cell type)"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid5 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1323912-1325831, 1471368-1472903 and
                     1492104-1493255). Subtraction by Bento Soares and M.
                     Fatima Bonaldo.  "
BASE COUNT      247 a    177 c    134 g    258 t      2 others
ORIGIN
```

```
  Query Match             30.0%;  Score 695.8;  DB 9;  Length 818;
  Best Local Similarity   95.9%;  Pred. No. 3e-171;
  Matches  745;  Conservative    0;  Mismatches   29;  Indels    3;  Gaps    3;
```

```
Qy   1547 gctaccggaaatgcttggggaaaaaagggcg-agttaagaaagcccacaaatccaagacc 1605
          ||||||| ||||||| |||||||||||||||| | |||||| |||||| ||| |||||| | 
Db    818 GCTACCGAAAATGCTGGGGGAAAAAAGGGCGAATTTAAGNAAGCCCNCAATCCCAAGAAC 759

Qy   1606 cattgagaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttccaaa 1665
            |||||||||||||||||  ||||||||||||  | |  | ||| |||||||||| || 
Db    758 ATTTGAGAAGTGGGTGGGAATAAAGGTAAAATTTGAACCATTCCCCAAGTCATTTCCCAA 699

Qy   1666 cttgaaaacagaatca-gtgttaaattcattttattcttattaaggaaatactttgcata 1724
          || |||||||||||||| | || ||||||||||||||||||||||||||||||||||| |||| 
Db    698 CTGGAAAACAGAATCAGGGGTAAAATTCATTTTATTCTTATTAAGGAAATACTTGGCATA 639

Qy   1725 aattaatcagcccc-agagtgctttaaaaaattctcttaaataaaaatatagactcgct 1783
          |||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||| 
Db    638 AATTAATCAGCCCCAAGAGTGCTTTTAAAAATTCTCTTAAATAAAAATAATAGACTCGCT 579

Qy   1784 agtcagtaaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgac 1843
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||| 
Db    578 AGTCAGTAAAGATATTTGAATATGTATCGTGCCCCCTCTGGTGTCTTTGATCAGGATGAC 519

Qy   1844 atgtgccatttttcagaggacgtgcagacaggctggcattctagattacttttcttactc 1903
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db    518 ATGTGCCATTTTTCAGAGGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTC 459

Qy   1904 tgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctactg 1963
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||| 
Db    458 TGAAACATGGCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCACGGCTGCCCCTACTG 399

Qy   1964 caaatggcagttttaatcttatcttttggcttctgcagatggttgcaattgatccttaac 2023
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||| 
Db    398 CAAATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTGGCAATTGATCCTTAAC 339

Qy   2024 caataatggtcagtcctcatctctgtcctgcttcataggtgccaccttgtgtgtttaaag 2083
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||| 
Db    338 CAATAATGGTCAGTCCTCATCTCTGTCGTGCTTCATAGGTGCCACCTTGTGTGTTTAAAG 279

Qy   2084 aagggaagctttgtacctttagagtgtaggtgaaatgaatgaatggcttggagtgcactg 2143
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db    278 AAGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTG 219

Qy   2144 agaacagcatatgatttcttgctttggggaaaaagaatgatgctatgaaattggtgggtg 2203
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db    218 AGAACAGCATATGATTTCTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTG 159

Qy   2204 gtgtatttgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaaaccca 2263
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db    158 GTGTATTTGAGAAGATAATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCA 99

Qy   2264 aaatacagctatgaagtgctgggcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db     98 AAATACAGCTATGAAGTGCTGGGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 42
```

```
RESULT   9
BG289777
LOCUS       BG289777                881 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602385071F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4514129 5',
            mRNA sequence.
ACCESSION   BG289777
VERSION     BG289777.1  GI:13045907
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 881)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
```

```
Qy    181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    187 GTGGTTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGGAAGATCACTGAATTGCACAGTG 246

Qy    241 aagacttattcaacttcatatacccctggaggatctggaccgggagttcaaggcttttgcc 300
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    247 AAGACTTATTCAACTTCATATACCCTGGAGGATCTGGACCGGGAGTTCAAGGCTTTTGCC 306

Qy    301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    307 CATGCTCAATGGAAAGCACAAGTACGAAGTATATATTCTCTATTAATGGGTTCATACAAT 366

Qy    361 gacatttttgactta-ttttttttcaaattgcaggagtttgtttaaagacaaaaaattagt 419
          ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    367 GACATTTTTGACTTATTTTTTTTTCAAATTGCAGGAGTTTGTTTAAAGACAAAAAATTAGT 426

Qy    420 agaatacttaaaggagagttc-ttttgatgcagtgtttctcgatccttttgataactgtg 478
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    427 AGAATACTTAAAGGAGAGTTCTTTTTGATGCAGTGTTTCTCGATCCTTTTGATAACTGTG 486

Qy    479 gcttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatacttt 538
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    487 GCTTAATTGTTGCCAAATATTTCTCCCTCCCCTCCGTGGTCTTCGCCAGGCGAATACTTT 546

Qy    539 gcccctatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattc 598
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    547 GCCACTATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTC 606

Qy    599 tcttagggttctcagatgccatgactttcaaggagagagtacggaaccac-atcatgcac 657
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    607 TCTTAGGGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTACGGAACCACAATCATGCAC 666

Qy    658 ttggaggaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaa 717
          |||||||||||||||||||||||| || |||||||||||||||||||| |||||||||||
Db    667 TTGGAGGAACATTTATTATGCCACCGATTATTCAAAAATGCCCTAGAAATAGGCTCTGAA 726

Qy    718 attctccaaacacctgtt-acggagtatga---tctctacagccacacatcaatttggtt 773
          |||||||||| ||||||| |||||||||||   ||| | |  |||||   |||| |  |
Db    727 ATTCTCCAAAGACCTGTTAACGGAGTATGAATCTCTACAGACACACACTCAAATTGGAAT 786

Qy    774 gttgcgaacgg--actttgttttggactatcccaaacccgt-gatgcccaacatgatctt 830
          ||||||||||   ||| |||  | |||||||| |||||||| ||||||||||||||||||
Db    787 GTTGCGAACGGGACCTTGGTTGGGAACTATCCAAAACCCGTGGATGCCCAACATGATCTT 846

Qy    831 cattggtggtatcaactgccatcagggaaa 860
          ||| || |  |  |    |  || | ||||
Db    847 CATGGGAGAAACAACTGTCATTCCGTGAAA 876
```

```
RESULT   7
BI762347
LOCUS       BI762347                884 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603049004F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5189040 5',
            mRNA sequence.
ACCESSION   BI762347
VERSION     BI762347.1  GI:15753925
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 884)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11472  row: h  column: 01
            High quality sequence stop: 796.
FEATURES             Location/Qualifiers
     source          1. .884
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5189040"
                     /clone_lib="NIH_MGC_116"
                     /lab_host="DH10B"
                     /note="Organ: pooled colon, kidney, stomach; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 3 colons, age 26 yo male, 49 yo
                     female, 71 yo male colon; 46 yo male kidney, and pool of 2
                     stomachs, 62 yo male and 70 yo female.  Library is
                     oligo-dT primed and  directionally cloned (EcoRV site is
                     destroyed upon cloning).  Average insert size 1.4 kb,
                     insert size range 1-3 kb.  Library is normalized and
                     enriched for full-length clones and was constructed by C.
                     Gruber  (Invitrogen).  Research Genetics tracking code
                     023. Note: this is a NIH_MGC Library."
BASE COUNT      213 a    207 c    210 g    253 t      1 others
ORIGIN
```

```
  Query Match           30.5%;  Score 707.4;  DB 10;  Length 884;
  Best Local Similarity   94.0%;  Pred. No. 2.8e-174;
  Matches  812;  Conservative    0;  Mismatches   42;  Indels   10;  Gaps    7;

Qy      26 gcacagggtggaccagccccctcctctatgtgtgtgtctgctgctgacctgtggctttg 85
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCACAGGGTGGACCAGCCCCCTTCCTCTATGTGTGTGTCTGCTGCTGACCTGTGGCTTTG 60

Qy      86 ccgaggcagggaagctactggtagtgcccatggatgggagccactggttcaccatgaggt 145
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCGAGGCAGGGAAGCTACTGGTAGTGCCCATGGATGGGAGCCACTGGTTCACCATGAGGT 120

Qy     146 cggtggtggagaaactcattctcaggggcatgaggtggttgtagtcatgccagaggtga 205
           |||||||.||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGGTGGTGGAGAAACTCATTCTCAGGGGGCATGAGGTGGTTGTAGTCATGCCAGAGGTGA 180

Qy     206 gttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttcatataccc 265
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GTTGGCAACTGGGAAGATCACTGAATTGCACAGTGAAGACTTATTCAACTTCATATACCC 240

Qy     266 tggaggatctggaccgggagttcaaggcttttgcccatgctcaatggaaagcacaagtac 325
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGGAGGATCTGGACCGGGAGTTCAAGGCTTTTGCCCATGCTCAATGGAAAGCACAAGTAC 300

Qy     326 gaagtatatattctctattaatgggttcatacaatgacatttttgacttatttttttcaa 385
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GAAGTATATATTCTCTATTAATGGGTTCATACAATGACATTTTTGACTTATTTTTTTCAA 360

Qy     386 attgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagagttcttttg 445
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATTGCAGGAGTTTGTTTAAAGACAAAAAATTAGTAGAATACTTAAAGGAGAGTTCTTTTG 420

Qy     446 atgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaatatttctccc 505
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ATGCAGTGTTTCTCGATCCTTTTGATAACTGTGGCTTAATTGTTGCCAAATATTTCTCCC 480

Qy     506 tcccctccctggtcttcgccaggggaatactttgccactatcttgaagaaggtgcacagt 565
           |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db     481 TCCCCTCCGTGGTCTTCGCCA-GGGAATACTTTGCCACTATCTTGAAGAAGGTGCACAGT 539

Qy     566 gccctgctcctctttcctatgtccccagaattctcttagggttctcagatgccatgac-- 623
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 GCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTCTTAGGGTTCTCAGATGCCATGACCT 599

Qy     624 tttcaaggagagagtacggaaccacatcatgcacttgg-aggaacatttattatgccacc 682
           ||  | |   |||||||||||||| |||||||||||||| ||||||||||||||||||||
Db     600 TTCAAGGAGCAGAGTACGGAACCTCATCATGCACTTGGCAGGAACATTTATTATGCCACC 659
```

found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10407 row: m column: 12
High quality sequence stop: 746.

```
FEATURES             Location/Qualifiers
     source          1. .882
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4516403"
                     /clone_lib="NIH_MGC_93"
                     /tissue_type="transitional cell papilloma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      259 a    155 c    202 g    266 t
ORIGIN
```

Query Match 31.1%; Score 722.4; DB 10; Length 882;
Best Local Similarity 94.5%; Pred. No. 3.3e-178;
Matches 792; Conservative 0; Mismatches 41; Indels 5; Gaps 4;

```
Qy   1447 gacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcgtg 1506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGTG 60

Qy   1507 ctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttgggg 1566
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGG 120

Qy   1567 aaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtgggaaa 1626
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGAAA 180

Qy   1627 taaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgtt 1686
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTGTT 240

Qy   1687 aaattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtgct 1746
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTGCT 300

Qy   1747 ttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaatat 1806
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAATAT 360

Qy   1807 gtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggacgt 1866
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGACGT 420

Qy   1867 gcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttgggagt 1926
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGAGT 480

Qy   1927 gcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatcttatc 1986
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTATC 540

Qy   1987 ttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatctc 2046
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCTCATCTC 600

Qy   2047 tgtcctgcttcataggtgccaccttgtgt-gtttaaagaagggaagctttgtacctttag 2105
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    601 TGTCCTGCTTCATAGGTGCCACCTTGTGTGGTTTAAAGAAGGGAAGCTTTGTACCTTTAG 660

Qy   2106 agtgtaggtg-aaatgaatgaatggcttggagtgcactgagaacagcatatgatttct-t 2163
          |||||||||| |||||||||||||| | |  ||||||||||||||||||||||||||| |
Db    661 AGTGTAGGTGAAAATGAATGAATGGTTGGAGGTGCACTGAGAACAGCATATGATTTCTGG 720

Qy   2164 gctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatc 2223
          |||||||||||||||||||||||||||| || | | ||| |     |||   | |  |||
Db    721 GCTTTGGGGAAAAAGAATGATGCTATGGAAATTGGGGGGTGGTGGATTGGAGAAAATATC 780

Qy   2224 attgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtg 2281
          ||  |||||||||| | |    |||||||||||||| || | |    |   | |||
Db    781 ATGCTTTATGTCAAAGGAGG--TGATTTGATAAAAACCAAATACAGGTATTAAAGGTG 836
```

```
RESULT   6
BI759297
LOCUS       BI759297                876 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603042939F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5183433 5',
            mRNA sequence.
ACCESSION   BI759297
VERSION     BI759297.1  GI:15750875
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 876)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11457  row: n  column: 10
            High quality sequence stop: 735.
FEATURES             Location/Qualifiers
     source          1. .876
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5183433"
                     /clone_lib="NIH_MGC_116"
                     /lab_host="DH10B"
                     /note="Organ: pooled colon, kidney, stomach; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 3 colons, age 26 yo male, 49 yo
                     female, 71 yo male colon; 46 yo male kidney, and pool of 2
                     stomachs, 62 yo male and 70 yo female.  Library is
                     oligo-dT primed and  directionally cloned (EcoRV site is
                     destroyed upon cloning).  Average insert size 1.4 kb,
                     insert size range 1-3 kb.  Library is normalized and
                     enriched for full-length clones and was constructed by C.
                     Gruber  (Invitrogen).  Research Genetics tracking code
                     023. Note: this is a NIH_MGC Library."
BASE COUNT      226 a    193 c    205 g    252 t
ORIGIN
```

Query Match 31.1%; Score 720.8; DB 10; Length 876;
Best Local Similarity 94.6%; Pred. No. 8.6e-178;
Matches 823; Conservative 0; Mismatches 37; Indels 10; Gaps 7;

```
Qy      1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7 CTCAGCTGCAGTTCTCTGATGGCTTGCACAGGGTGGACCAGCCCCCTTCCTCTATGTGTG 66

Qy     61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     67 TGTCTGCTGCTGACCTGTGGCTTTGCCGAGGCAGGCAAGCTACTGGTAGTGCCCATGGAT 126

Qy    121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggggcatgag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    127 GGGAGCCACTGGTTCACCATGAGGTCGGTGGTGGAGAAACTCATTCTCAGGGGGCATGAG 186
```

```
Qy   1676 gaatcagtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagc 1735
          ||  |||||||||| ||| || || || ||| ||||||| ||  |   |  | ||
Db   1689 -AACAAGTGTTAAATCCATATTGTTTTTGTTAGGGAAATAATTCACCATACATTATACAT 1747

Qy   1736 cccagagtgctttaaaaaattctctta--aataaaaataatagactcgctagtcagtaaa 1793
          |   |    | ||| |||  |  ||  |||  || |  | | | | |   |||| ||
Db   1748 TCAGCACATTTAAAAATAATAATAATAACAATCTAATTGCTGGCCACACCCATCAGGGAA 1807

Qy   1794 gatatttgaatatgtatcgtgcccccctccggtgtctttgatcaggatgacatgtgccatt 1853
          | | | | ||||||   ||| |   ||| || ||||   ||  ||  ||   ||||||
Db   1808 GGTTCTAGTATATGTGATGTG-CTTTTCCAGTATCTTCAGTCTAGACAACTCTTGCCATC 1866

Qy   1854 tttcagaggacgtgcagacaggctggcattctagatt 1890
          | |  |   |  | |||| || |||||| |||   ||
Db   1867 TGT-TGGTAATTTACAGAAAGTCTGGCACTCTGCTTT 1902


RESULT   4
BG402908
LOCUS       BG402908                 868 bp    mRNA    linear   EST 12-MAR-2001
DEFINITION  602418710F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4525755 5',
            mRNA sequence.
ACCESSION   BG402908
VERSION     BG402908.1  GI:13296356
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 868)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10432  row: c  column: 04
            High quality sequence stop: 785.
FEATURES             Location/Qualifiers
     source          1. .868
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4525755"
                     /clone_lib="NIH_MGC_93"
                     /tissue_type="transitional cell papilloma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      258 a    151 c    187 g    272 t
ORIGIN


  Query Match             31.9%;  Score 741.2;  DB 10;  Length 868;
  Best Local Similarity   95.9%;  Pred. No. 3.8e-183;
  Matches  804;  Conservative    0;  Mismatches   28;  Indels    6;  Gaps    4;

Qy   1487 gtttcctcttggccgtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatg 1546
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTTTCCTCTGGGCCGTCGTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATG 60

Qy   1547 gctaccggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaagaccc 1606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCTACCGGAAATGCTTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCC 120

Qy   1607 attgagaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttccaaac 1666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTGAGAAGTGGGTGGGAAATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAAC 180

Qy   1667 ttgaaaacagaatcagtgttaaattcattttattcttattaaggaaatactttgcataaa 1726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTGAAAACAGAATCAGTGTTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAA 240

Qy   1727 ttaatcagcccagagtgctttaaaaaattctcttaaataaaaataatagactcgctagt 1786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTAATCAGCCCAGAGTGCTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGT 300

Qy   1787 cagtaaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgacatg 1846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CAGTAAAGATATTTGAATATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATG 360

Qy   1847 tgccatttttcagaggacgtgcagacaggctggcattctagattacttttcttactctga 1906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGCCATTTTTCAGAGGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGA 420

Qy   1907 aacatggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaa 1966
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AACATGGCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAA 480

Qy   1967 atggcagttttaatcttatcttttggcttctgcagatggttgcaattgatccttaaccaa 2026
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAA 540

Qy   2027 taatggtcagtcctcatctctgtcctgcttcataggtgccaccttgtgtgtttaaagaag 2086
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TAATGGTCAGTCCTCATCTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAG 600

Qy   2087 ggaagctttgtaccttagagtgtaggtgaaatgaatgaatggcttggagtgcactgaga 2146
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATGGTTTGGAGTGCACTGAGA 660

Qy   2147 acag-catatga-tttcttgctttggggaaaaagaatgatgctatgaaattggtgggtgg 2204
          |||| |||||| ||||||||||||||||||||||||||||||||| ||||||||||
Db    661 ACAGACATATGATTTTCTTGCTTTGGGGAAAAAGAATGATGCTATGAAAATGGTGGGTGG 720

Qy   2205 tgta--tttgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaaaccc 2262
          ||||  ||    |||||||||  |  ||||||| ||  | | ||||||||||||||||||
Db    721 TGTAATTTGAGAAAGATAATCCATTGTTATGTCCAAATGGGGTGAATTTGATAAAAACCC 780

Qy   2263 aaaatacagctatgaa--gtgctgggcaagtttacttttttttctgatgtttcctacaa 2318
          ||||      ||||||      |||||||||| |||||||| ||||||||||||||||
Db    781 AAAAATACAGTATGAAAGTGCTTGGGCAAGTTAACTTTTTTCCTGATGTTTCCTACAA 838


RESULT   5
BG291462
LOCUS       BG291462                 882 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602387347F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4516403 5',
            mRNA sequence.
ACCESSION   BG291462
VERSION     BG291462.1  GI:13049372
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 882)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
```

```
                     /db_xref="MGD:MGI:1910585"
                     /db_xref="taxon:10090"
                     /clone="0610013P07"
                     /sex="male"
                     /tissue_type="kidney"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     gene            28. .1629
                     /gene="Ugt1a1"
     CDS             28. .1629
                     /gene="Ugt1a1"
                     /note="UDP-glucuronosyltransferase 1 family, member 1
                     data source:MGD, source key:MGI:98898, evidence:ISS
                     putative"
                     /codon_start=1
                     /protein_id="BAB22243.1"
                     /db_xref="GI:12832754"
                     /translation="MDTGLCVPLRGISGLLLLLCALPWAEGAKVLVLPMEGSQWLSMR
                     DVVRELHARGHQTVVLASEVTVHIKGEDFFTLKTYAFPYTKEEYQQEILSDIEKTFKT
                     QHFVKAFFETTASIRNFFDLYSNSCIALLHNKMLIQQLNSSFFDVILTDPIFPCGAVL
                     AKYLQIPAVFILRSLSCGIEYEATQCPNPSSYIPNLLTRLSDHMDFLQRVQNMLYYLV
                     LKYICRLSITPYESLASELLQREVSLVEVLSHASVWLFRGDFVLDYPRPIMPNMVFIG
                     GINCVTKKPLSQEFEAYVNASGEHGIVVFSLGSMVSEIPEKKAMEIAEALGRIPQTVL
                     WRYTGTRPSNLAKNTILVKWLPQNDLLGHPKTRAFITHSGSHGIYEGICNGVPMVMMP
                     LFGDQMDNAKRMETRGAGVTLNVLEMTADDLENALKTVINNKSYKENIMRLSSLHKDR
                     PIEPLDLAVFWVEYVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAIVLTVVFIVFK
                     CCAYGCRKCFGGKGRVKKSHKSKTH"
     polyA_signal    2183. .2188
                     /note="putative"
     polyA_site      2202
                     /note="putative"
BASE COUNT      568 a    514 c    514 g    606 t
ORIGIN
```

```
  Query Match             33.6%;  Score 780.2;  DB 11;  Length 2202;
  Best Local Similarity   65.9%;  Pred. No. 3.1e-193;
  Matches 1211;  Conservative    0;  Mismatches  613;  Indels   13;  Gaps    5;

Qy      59 tgtgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatgg 118
           ||   ||||||   ||     ||  || || |  |  ||  | |||||  | |||||||
Db      74 TGCTCCTGCTGTGTGCCCTGCCCTGGGCTGAAGGTGCAAAAGTGCTGGTGTTACCCATGG 133

Qy     119 atgggagccactggttcaccatgaggtcggtggtggagaaactcattctcagggggcatg 178
           | || ||||| ||| | | |||||||   || |||  | | |||     | | || ||
Db     134 AGGGCAGCCAGTGGCTGAGCATGAGGGATGTTGTGCGGGAGCTCCACGCCCGAGGTCACC 193

Qy     179 aggtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacag 238
           ||   || ||  |    |||||||||  |  |||  |   | ||  |       | |||
Db     194 AGACTGTGGTCCTGGCTTCAGAGGTGACTGTGCACATCAAAGGAGAAGACTTCTTCACCC 253

Qy     239 tgaagacttattcaacttcatataccctggagg---atctggaccgggagttcaaggctt 295
           | || || ||| |   | |||||||||  ||| |   ||| | |         | |  |
Db     254 TCAAAACCTATGCCTTTCCATATACCAAGGAAGAATATCAGCAAGAAATACTGAGCGACA 313

Qy     296 ttgcccatgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcat 355
           |||   |  |       | |  |||  |     |     | || |  |   |  |  ||
Db     314 TTGAGAAGACCTTTAAAACACAACATTTTGTGAAGGCCTTTTTTGAAACTACAGCATCTA 373

Qy     356 acaatgacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaat 415
             |   || |||||||| | |  |       ||| |    ||| ||  |  | || |
Db     374 TAAGAAACTTTTTTGACCTCTACTCAAATTCTTGTATAGCTTTATTGCACAATAAGATGC 433

Qy     416 tagtagaatacttaaaggagagttctttttgatgcagtgtttctcgatccttttgataact 475
           |  |  |  |  | ||    |||| ||||||||   | ||    || ||  ||     ||
Db     434 TGATCCAGCAACTGAATTCCAGTTTTTTTGATGTGATCTTAACAGACCCCATTTTCCCCT 493

Qy     476 gtggcttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatac 535
           ||||   | |  | ||||| ||  |     | ||  | ||| |  |     |      |
Db     494 GTGGAGCAGTACTGGCCAAGTACCTACAGATTCCTGCTGTGTTTATTCTGCGTTCTCTTT 553

Qy     536 tttgccactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaa 595
             ||  |       | || |   |||| || ||   ||||   || ||| | || |
```

```
Db     554 CCTGTGGCATAGAATATGAGGCCACACAATGTCCAAATCCTTCCTCTTATATTCCGAACC 613

Qy     596 ttctcttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgc 655
           | |||  | || | || ||   ||||   |||  | | || || | ||||    |
Db     614 TACTCACAAGGCTTTCTGACCACATGGACTTCCTGCAAAGGGTCCAGAACATGCTGTACT 673

Qy     656 acttggaggaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctg 715
           |  |||       | |   | ||||     |   |||          |    | |||||||
Db     674 ATCTGGTCCTGAAGTACATTTGCCGTTTATCAATCACTCCCTATGAAAGCCTGGCCTCTG 733

Qy     716 aaattctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgt 775
           |  || | || | |   ||  |       ||  |   ||||||  ||||  | ||| |||
Db     734 AGCTTTTGCAGAGAGAAGTATCCTTAGTGGAGGTTCTCAGCCATGCATCCGTGTGGCTGT 793

Qy     776 tgcgaacggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattg 835
           | |||  |||||||||  | ||||| ||||  ||| | ||||| |||||| |||||||||
Db     794 TCCGAGGGGACTTTGTACTCGACTACCCCAGGCCCATCATGCCTAACATGGTCTTCATTG 853

Qy     836 gtggtatcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatg 895
           | || || |||||    |       |||||  |  ||  |||||||||||||||  | || |
Db     854 GGGGCATAAACTGTGTTACCAAGAAGCCCCTCTCTCAGGAATTTGAAGCCTATGTCAACG 913

Qy     896 cttctggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccag 955
           | ||||| || ||||| || || |||||||||||||||| ||||||||||| ||||| |
Db     914 CCTCTGGGGAGCATGGCATCGTTGTTTTCTCTTTGGGATCCATGGTCTCAGAGATTCCGG 973

Qy     956 agaagaaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggc 1015
           |||||||||| |||| |||||||||| |||||||||| ||| ||||||| ||||||||||
Db     974 AGAAGAAAGCCATGGAAATTGCTGAGGCTTTGGGCAGAATTCCTCAGACGGTCCTGTGGC 1033

Qy    1016 ggtacactggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctac 1075
           | ||||| |||||  |||||||||||||||| || ||||| || ||||| || |||||||
Db    1034 GCTACACCGGAACTAGACCATCGAATCTTGCAAAGAACACAATTCTTGTCAAATGGCTAC 1093

Qy    1076 cccaaaacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttccc 1135
           ||||||| |||||||||||||| || | ||| || || || ||||| ||  |||| ||||
Db    1094 CCCAAAATGATCTGCTTGGTCATCCAAAGACTCGGGCATTCATCACACACTCTGGCTCCC 1153

Qy    1136 atggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgccccttgtttggtg 1195
           ||||| |||||||| | |||||||||| ||||| |||||||||||||| | ||||| |
Db    1154 ATGGTATTTATGAAGGAATATGCAATGGAGTTCCGATGGTGATGATGCCCCTATTTGGCG 1213

Qy    1196 atcagatggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttc 1255
           |||||||||||||||| |||||||||| |||  |||||||| ||||||||||||||| |
Db    1214 ATCAGATGGACAATGCCAAGCGCATGGAAACTCGGGGAGCTGGGGTGACCCTGAATGTCC 1273

Qy    1256 tggaaatgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagtt 1315
           | ||||||||| |||| ||||| |||||||| || ||| | |||||||  |||| || |
Db    1274 TTGAAATGACTGCTGATGATTTGGAAAATGCCCTTAAAACTGTCATCAACAACAAGAGCT 1333

Qy    1316 acaaggagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctgg 1375
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||  | ||||| ||||
Db    1334 ACAAGGAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGTCCTATAGAGCCTCTGG 1393

Qy    1376 acctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgcc 1435
           |||||||  |||||||||||||||| |  |||||||||||||||| || ||||||||||||
Db    1394 ACCTGGCTGTGTTCTGGGTGGAATACGTGATGAGGCACAAGGGGGCACCACACCTGCGCC 1453

Qy    1436 ccgcagcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctct 1495
           | || ||||| |||||||||||||| |||||||| |||||||| |||||||| ||||||
Db    1454 CGGCCGCCCATGACCTCACCTGGTATCAGTACCACTCCTTGGATGTGATTGGCTTCCTCC 1513

Qy    1496 tggccgtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccgga 1555
           ||||| | ||| |||||||||| ||||||   |||||||||||||||| ||||||| |||||
Db    1514 TGGCCATTGTGTTGACAGTGGTCTTCATTGTCTTTAAATGTTGTGCCTATGGCTGCCGGA 1573

Qy    1556 aatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaag 1615
           ||||||| |||  ||| |||||||| |||||| | ||||||||||||||||||||||||||
Db    1574 AATGCTTTGGGGGAAAGGGGCGAGTGAAGAAATCACACAAATCCAAGACCCATTGAGAAG 1633

Qy    1616 tgggtgggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaaca 1675
           |||| || |   ||||  || | ||    |||| ||  ||| ||  ||  | ||
Db    1634 TGGGGGGAAGTGAAGGAGAAGTATTAGTTCATTATCTGATCAGTTGAAACTTGGA----- 1688
```

```
Qy   1073 taccccaaaacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggtt 1132
          |||||||||| |||||||||||||| || | ||| || || || ||||| ||  |||| |
Db   1180 TACCCCAAAATGATCTGCTTGGTCATCCAAAGACTCGGGCATTCATCACACACTCTGGCT 1239

Qy   1133 cccatggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttg 1192
          |||||||| |||||||| | |||||||||||| ||||| |||||||||||||||| | ||||
Db   1240 CCCATGGTATTTATGAAGGAATATGCAATGGAGTTCCGATGGTGATGATGCCCCTATTTG 1299

Qy   1193 gtgatcagatggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatg 1252
          | |||||||||||||||| |||||||||| ||| |||||||| ||||||||||||||
Db   1300 GCGATCAGATGGACAATGCCAAGCGCATGGAAACTCGGGGAGCTGGGGTGACCCTGAATG 1359

Qy   1253 ttctggaaatgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaa 1312
          | || ||||||||| |||| ||||| |||||||| || ||| | ||||||||  |||| |
Db   1360 TCCTTGAAATGACTGCTGATGATTTGGAAAATGCCCTTAAAACTGTCATCAACAACAAGA 1419

Qy   1313 gttacaaggagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgc 1372
          | |||||||||||||||||||||||||||||||||||||||||||||||||| ||  | ||||| |
Db   1420 GCTACAAGGAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGTCCTATAGAGCCTC 1479

Qy   1373 tggacctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgc 1432
          |||||||||| |||||||||||||| |  |||||||||||||||| || ||||||||||
Db   1480 TGGACCTGGCTGTGTTCTGGGTGGAATACGTGATGAGGCACAAGGGGGCACCACACCTGC 1539

Qy   1433 gccccgcagcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcc 1492
          |||| || ||||| |||||||||||||| |||||||| |||||||| |||||||| ||||
Db   1540 GCCCGGCCGCCCATGACCTCACCTGGTATCAGTACCACTCCTTGGATGTGATTGGCTTCC 1599

Qy   1493 tcttggccgtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctacc 1552
          || ||||| | ||| |||||||||| ||||||   |||||||||||||||| ||||||| ||
Db   1600 TCCTGGCCATTGTGTTGACAGTGGTCTTCATTGTCTTTAAATGTTGTGCCTATGGCTGCC 1659

Qy   1553 ggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgag 1612
          |||||||||| |||  ||| |||||||| |||||| | ||||||||||||||||||||
Db   1660 GGAAATGCTTTGGGGGAAAGGGGCGAGTGAAGAAATCACACAAATCCAAGACCCATTGAG 1719

Qy   1613 aagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaa 1672
          ||||||| || |   ||||  || | ||    |||| ||  ||| ||  ||  | ||
Db   1720 AAGTGGGGGGAAGTGAAGGAGAAGTATTAGTTCATTATCTGATCAGTTGAAACTTGGA-- 1777

Qy   1673 acagaatcagtgttaaattcattttattcttattaaggaaatactttgcataaattaatc 1732
               ||  |||||||  || ||| || || || ||| ||||||| ||  |   |  | ||
Db   1778 ----AACAAGTGTTAGATCCATATTGTTTTTGTTAGGGAAATAATTCACCATACATTATA 1833

Qy   1733 agccccagagtgctttaaaaaattctctta--aataaaaataatagactcgctagtcagt 1790
               |   |     |  ||| |||  |  ||  |||  || |  | | | |  ||||
Db   1834 CATTCAGCACATTTAAAAATAATAATAATAACAATCTAATTGCTGGCCACACCCATCAGG 1893

Qy   1791 aaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgcc 1850
           ||| |  | | ||||||   ||| |   ||| || |||||   ||  ||  ||  ||||
Db   1894 GAACGTTCTAGTATATGTGATGTG-CTTTTCCAGTATCTTCAGTCTAGACAACTCTTGCC 1952

Qy   1851 atttttcagaggacgtgcagacaggctggcattctagatt 1890
          || | |  |   | | |||| || |||||| |||   ||
Db   1953 ATCTGT-TGGTAATTTACAGAAAGTCTGGCACTCTGCTTT 1991
```

RESULT 3
AK002629

```
LOCUS       AK002629               2202 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus adult male kidney cDNA, RIKEN full-length enriched
            library, clone:0610013P07:UDP-glucuronosyltransferase 1 family,
            member 1, full insert sequence.
ACCESSION   AK002629
VERSION     AK002629.1  GI:12832753
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:0610013P07.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 2202)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schrml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0. Second strand cDNA
            was prepared with the primer adapter of sequence[5'
            GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was
            cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
            XhoI. Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .2202
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
```

```
Db   2048 AAATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACC 2107

Qy   2025 aataatggtcagtcctcatctctgtcctgcttcataggtgccaccttgtgtgtttaaaga 2084
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2108 AATAATGGTCAGTCCTCATCTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGA 2167

Qy   2085 agggaagctttgtacctttagagtgtaggtgaaatgaatgaatggcttggagtgcactga 2144
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2168 AGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGA 2227

Qy   2145 gaacagcatatgatttcttgctttggggaaaaagaatgatgctatgaaattggtgggtgg 2204
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2228 GAACAGCATATGATTTCTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGG 2287

Qy   2205 tgtatttgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaacccaa 2264
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2288 TGTATTTGAGAAGATAATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAA 2347

Qy   2265 aatacagctatgaagtgctgggcaagtttacttttttttctgatgtttcctacaact 2320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2348 AATACAGCTATGAAGTGCTGGGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 2403


RESULT   2
BC012716
LOCUS       BC012716               2306 bp    mRNA    linear   HTC 20-AUG-2001
DEFINITION  Mus musculus, Similar to UDP-glucuronosyltransferase 1 family,
            member 1, clone IMAGE:3993384, mRNA.
ACCESSION   BC012716
VERSION     BC012716.1  GI:15215247
KEYWORDS    HTC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 18 Row: e Column: 4
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 801898
            This clone has the following problem: incomplete processing.
FEATURES             Location/Qualifiers
     source          1. .2306
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3993384"
                     /tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
                     old, gross tissue."
                     /clone_lib="NCI_CGAP_Mam5"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
BASE COUNT      594 a    559 c    528 g    625 t
ORIGIN
```

Query Match 34.0%; Score 789.6; DB 11; Length 2306;
Best Local Similarity 66.6%; Pred. No. 1.1e-195;
Matches 1225; Conservative 0; Mismatches 599; Indels 16; Gaps 6;

```
Qy     56 gtgtgtgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgccca 115
          | || | ||| |      | || | ||  |     |  |  ||||| ||||| || |||
Db    163 GGGTTTCTCTTCCTAGTGCTTGGGCCTCAGTTCTAGGTGACAAGCTGCTGGTGGTCCCCC 222

Qy    116 tggatgggagccactggttcaccatgaggtcggtggtggagaaactcattctcagggggc 175
           |||||| ||||||||| | | ||||| |  | | |||||  | |||| |    | || |
Db    223 AGGATGGAAGCCACTGGCTTAGCATGAAGGAGATAGTGGAACATCTCAGTGAACGAGGAC 282

Qy    176 atgaggtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgca 235
          ||||  |  | ||  |  ||||||| || | || ||   |||||  |||     | | ||
Db    283 ATGACATTATGGTCCTAGTGCCAGAAGTCAATTTGCTTTTGGGAGAATCCAAATACTACA 342

Qy    236 cagtgaagacttattcaacttcatataccctggaggatctggaccgggagttcaaggctt 295
            ||| |  |  ||| | | || | ||| ||||| ||  |       ||     | |
Db    343 GGAGGAAAATCTTCTCAGTTACCTACAGCCTAGAGGAACTACAGACCCGTTTTCGCACCT 402

Qy    296 ttgcccatgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcat 355
          |||   |    || |      |    |        ||   | |||||  |  |
Db    403 TTGGAAACAACCACTTTCTTCCCAGTGCCGCCCTGATTGGTCCTCTAAGAGAGTACAGGA 462

Qy    356 acaatgacatt---tttgacttatttttttcaaattgcaggagtttgtttaaagacaaaa 412
          ||||    |||    | ||| | |  |||| |  |||  |||  |  | || |||
Db    463 ACAACATGATTGTCGTGGACATGTGCTTTTTCAGCTGCCAGAGCCTCCTGAAGGACTCTG 522

Qy    413 aattagtagaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgata 472
              |  | | | | |||||     ||||||||  ||||    || ||
Db    523 CCACCCTCAGCTTCCTCAGGGAGAACAAGTTTGATGGTCTGTTCACAGACCCAGCCATGC 582

Qy    473 actgtggcttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaa 532
           |||||  | ||  | ||  | ||| ||  ||| ||||| ||  |  ||  ||| ||
Db    583 CCTGTGGTGTGATCCTGGCTGAGTATCTCAACCTGCCCTCTGTCTACCTCTTCAGAGGTT 642

Qy    533 tactttgccactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtcccca 592
          | |  |||          |        ||  ||  ||||    ||| | ||||| || ||||
Db    643 TCCCATGCTCTCTGGAACACATGCTTGGTCAAAGCCCAAGCCCTGTGTCCTACGTTCCCA 702

Qy    593 gaattctcttagggttctcagatgccatgactttcaaggagagagtacggaaccacatca 652
          || |   |      ||||||||   |||||| ||      |  |  |   |||   ||
Db    703 GATTCTACACCAAATTCTCAGACCACATGACATTTCCCCAACGGCTGGCCAACTTTATTG 762

Qy    653 tgcacttggaggaacatttattatgccaccgttttttcaaaaatgccctagaaatagcct 712
          |  || |  |||| | |   |||   |  || | |    |||      |   || ||||
Db    763 TTAACATCTTGGAAAACTACCTATATTATTGTCTGTATTCAAAGTATGAAATCATTGCCT 822

Qy    713 ctgaaattctccaaacacctgttacggagtatgatctctacagccacacatcaatttggt 772
          | ||  | ||| | | |  |||  |      |    |  |   ||| |  ||  | |||
Db    823 CAGACCTCCTCAAGAGAGATGTGTCCCTACCTTCCTTACA---CCAGAATTCTCTGTGGC 879

Qy    773 tgttgcgaacggactttgttttggactatcccaaacccgtgatgcccaacatgatcttca 832
          ||||  ||    || ||||| || || |||||||  || || ||||||||||||||||||
Db    880 TGTTACGGTACGATTTTGTGTTCGAATATCCCAGGCCAGTCATGCCCAACATGATCTTCA 939

Qy    833 ttggtggtatcaactgccatcagggaaagccgttgcctatggaatttgaagcctacatta 892
          |||| || ||||||||  |  ||        | || ||  ||||||||||||||  | |
Db    940 TTGGAGGGATCAACTGTAAGAAGAAGGGAAAGCTGACTCAGGAATTTGAAGCCTATGTCA 999

Qy    893 atgcttctggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattc 952
          | || ||||| || ||||| || ||||||||||||||||||| ||||||||||| ||||
Db   1000 ACGCCTCTGGGGAGCATGGCATCGTGGTTTTCTCTTTGGGATCCATGGTCTCAGAGATTC 1059

Qy    953 cagagaagaaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgt 1012
          | |||||||||| |||| |||||||| |||||||||| ||| |||||||| |||||||
Db   1060 CGGAGAAGAAAGCCATGGAAATTGCTGAGGCTTTGGGCAGAATTCCTCAGACGGTCCTGT 1119

Qy    1013 ggcggtacactggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggc 1072
           |||| ||||| |||||  |||||||||||||||| || ||||| || ||||| || ||||
Db   1120 GGCGCTACACCGGAACTAGACCATCGAATCTTGCAAAGAACACAATTCTTGTCAAATGGC 1179
```

```
     source          1. .2450
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3854370"
                     /tissue_type="Colon, adenocarcinoma"
                     /clone_lib="NIH_MGC_65"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
BASE COUNT      669 a    509 c    565 g    707 t
ORIGIN
```

Query Match 70.3%; Score 1631.6; DB 11; Length 2450;
Best Local Similarity 82.6%; Pred. No. 0;
Matches 1880; Conservative 0; Mismatches 394; Indels 2; Gaps 1;

```
Qy      47 ttcctctatgtgtgtgtctgctgctgacctgtggctttgccgaggcagggaagctactgg 106
           || ||  | | || |   |  |       || ||| | |  |  |  |  ||||| ||||
Db     128 TTTCTGCAGGGGTTTTCTTCTTAGCACTTTGGGGCATGGTTGTAGGTGACAAGCTGCTGG 187

Qy     107 tagtgcccatggatgggagccactggttcaccatgaggtcggtggtggagaaactcattc 166
           | || ||   ||| || |||||||| | |  |||| |    | || |||   |||| |
Db     188 TGGTCCCTCAGGACGGAAGCCACTGGCTTAGTATGAAGGATATAGTTGAGGTTCTCAGTG 247

Qy     167 tcaggggcatgaggtggttgtagtcatgccagaggtgagttggcaactgggaagatcac 226
           | |||| |||||| | || || ||  |||| || || | || ||   ||  |  |||
Db     248 ACCGGGGTCATGAGATTGTAGTGGTGGTGCCTGAAGTTAATTTGCTTTTGAAAGAATCCA 307

Qy     227 tgaattgcacagtgaagacttattcaacttcatataccctggaggatctggaccgggagt 286
           | | ||||   || |  ||| ||    | |||  ||  || || ||| |       |
Db     308 AATACTACACAAGAAAAATCTATCCAGTGCCGTATGACCAAGAAGAGCTGAAGAACCGTT 367

Qy     287 tcaaggcttttgcccatgctcaatgg--aaagcacaagtacgaagtatatattctctatt 344
           | |  | ||||   |   ||| | |   |   | | | || |     |     |
Db     368 ACCAATCATTTGGAAACAATCACTTGCTGAGCGATCATTCCTAACTGCTCCTCAGACAGA 427

Qy     345 aatgggttcatacaatgacatttttgacttatttttttcaaattgcaggagtttgtttaa 404
           |     |||      || ||| | | |  ||    || |||  |||  |  |  |
Db     428 GTACAGGAATAACATGATTGTTATTGGCCTGTACTTCATCAACTGCCAGAGCCTCCTGCA 487

Qy     405 agacaaaaaattagtagaatacttaaaggagagttcttttgatgcagtgtttctcgatcc 464
           ||||   |     |  | | ||| ||||||||    ||||||||  | ||    || ||
Db     488 GGACAGGGACACCCTGAACTTCTTTAAGGAGAGCAAGTTTGATGCTCTTTTCACAGACCC 547

Qy     465 ttttgataactgtggcttaattgttgccaaatatttctccctcccctccgtggtcttcgc 524
                 ||||||  | ||  | ||  | |||||   ||| || || |||  | ||
Db     548 AGCCTTACCCTGTGGGGTGATCCTGGCTGAGTATTTGGGCCTACCATCTGTGTACCTCTT 607

Qy     525 caggggaatactttgccactatcttgaagaaggtgcacagtgccctgctcctctttccta 584
           ||||||  | |  ||  |       |   |          |||| |  ||| | |||||
Db     608 CAGGGGTTTTCCGTGTTCCCTGGAGCATACATTCAGCAGAAGCCCAGACCCTGTGTCCTA 667

Qy     585 tgtccccagaattctcttagggttctcagatgccatgactttcaaggagagagtacggaa 644
           | |||||      |  | ||| |||||   |||||||||     |  ||||    ||
Db     668 CATTCCCAGGTGCTACACAAAGTTTTCAGACCACATGACTTTTTCCCAACGAGTGGCCAA 727

Qy     645 ccacatcatgcacttggaggaacatttattatgccaccgttttttcaaaaatgccctaga 704
           |  | |  |  | |||  ||| |  |   |||   |  || | ||   |||       |||
Db     728 CTTCCTTGTTAATTTGTTGGAGCCCTATCTATTTTATTGTCTGTTTTCAAAGTATGAAGA 787

Qy     705 aatagcctctgaaattctccaaacacctgttacggagtatgatctctacagccacacatc 764
           | | || || |   | ||| | | |  ||          | ||     |    ||
Db     788 ACTCGCATCAGCTGTCCTCAAGAGAGATGTGGATATAATCACCTTATATCAGAAGGTCTC 847

Qy     765 aatttggttgttgcgaacggactttgttttggactatcccaaacccgtgatgcccaacat 824
           |||||  ||||  ||   ||||||||  | || |||||  |  || || ||||||||||
Db     848 TGTTTGGCTGTTAAGATATGACTTTGTGCTTGAATATCCTAGGCCGGTCATGCCCAACAT 907

Qy     825 gatcttcattggtggtatcaactgccatcagggaaagccgttgcctatggaatttgaagc 884
           | |||||||||| ||||||||||  |  || | ||    ||| ||  ||||||||||||
Db     908 GGTCTTCATTGGAGGTATCAACTGTAAGAAGAGGAAAGACTTGTCTCAGGAATTTGAAGC 967

Qy     885 ctacattaatgcttctggagaacatggaattgtggttttctctttgggatcaatggtctc 944
```

```
                                                                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     968 CTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTC 1027

Qy     945 agaaattccagagaagaaagctatggcaattgctgatgctttgggcaaaatccctcagac 1004
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1028 AGAAATTCCAGAGAAGAAAGCTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGAC 1087

Qy    1005 agtcctgtggcggtacactggaacccgaccatcgaatcttgcgaacaacacgatacttgt 1064
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1088 AGTCCTGTGGCGGTACACTGGAACCCGACCATCGAATCTTGCGAACAACACGATACTTGT 1147

Qy    1065 taagtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcctttatcaccca 1124
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1148 TAAGTGGCTACCCCAAAACGATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCA 1207

Qy    1125 tgctggttcccatggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgcc 1184
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1208 TGCTGGTTCCCATGGTGTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCC 1267

Qy    1185 cttgtttggtgatcagatggacaatgcaaagcgcatggagactaagggagctggagtgac 1244
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1268 CTTGTTTGGTGATCAGATGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGAC 1327

Qy    1245 cctgaatgttctggaaatgacttctgaagatttagaaaaatgctctaaaagcagtcatcaa 1304
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1328 CCTGAATGTTCTGGAAATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAA 1387

Qy    1305 tgacaaaagttacaaggagaacatcatgcgcctctccagccttcacaaggaccgcccggt 1364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1388 TGACAAAAGTTACAAGGAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGT 1447

Qy    1365 ggagccgctggacctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgcc 1424
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1448 GGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCC 1507

Qy    1425 acacctgccccccgcagcccacgacctcacctggtaccagtaccattccttggacgtgat 1484
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1508 ACACCTGCCCCCCGCAGCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGAT 1567

Qy    1485 tggtttcctcttggccgtcgtgctgacagtggccttcatcacctttaaatgttgtgctta 1544
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1568 TGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTA 1627

Qy    1545 tggctacccgaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaagac 1604
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1628 TGGCTACCCGAAATGCTTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGAC 1687

Qy    1605 ccattgagaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttccaa 1664
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1688 CCATTGAGAAGTGGGTGGGAAATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAA 1747

Qy    1665 acttgaaaacagaatcagtgttaaattcattttattcttattaaggaaatactttgcata 1724
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1748 ACTTGAAAACAGAATCAGTGTTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATA 1807

Qy    1725 aattaatcagccccagagtgctttaaaaaattctcttaaataaaaataatagactcgcta 1784
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1808 AATTAATCAGCCCCAGAGTGCTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTA 1867

Qy    1785 gtcagtaaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgaca 1844
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1868 GTCAGTAAAGATATTTGAATATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACA 1927

Qy    1845 tgtgccatttttcagaggacgtgcagacaggctggcattctagattacttttcttactct 1904
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1928 TGTGCCATTTTTCAGAGGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCT 1987

Qy    1905 gaaacatgccctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctactgc 1964
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1988 GAAACATGCCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGC 2047

Qy    1965 aaatggcagttttaatcttatcttttggcttctgcagatggttgcaattgatccttaacc 2024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 21, 2002, 02:43:12 ; Search time 4821.79 Seconds
(without alignments)
6494.048 Million cell updates/sec

Title: US-09-835-082-1
Perfect score: 2320
Sequence: 1 ctcagctgcagttctctgat.....ttctgatgtttcctacaact 2320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1631.6	70.3	2450	11	BC011409	BC011409 Homo sapi
	2	789.6	34.0	2306	11	BC012716	BC012716 Mus muscu
	3	780.2	33.6	2202	11	AK002629	AK002629 Mus muscu
	4	741.2	31.9	868	10	BG402908	BG402908 602418710
	5	722.4	31.1	882	10	BG291462	BG291462 602387347
	6	720.8	31.1	876	10	BI759297	BI759297 603042939
	7	707.4	30.5	884	10	BI762347	BI762347 603049004
c	8	695.8	30.0	818	9	AI307269	AI307269 tb24d12.x
	9	695.4	30.0	881	10	BG289777	BG289777 602385071
	10	680.6	29.3	1005	10	BE870911	BE870911 601448862
	11	679.2	29.3	697	10	BG747143	BG747143 602704364
c	12	669.4	28.9	764	9	AI831562	AI831562 wj39b01.x
	13	644	27.8	723	10	BG424452	BG424452 602447772
	14	641.2	27.6	829	10	BG469506	BG469506 602532818
	15	640	27.6	833	10	BG400591	BG400591 602464244
c	16	639.8	27.6	791	9	AI765716	AI765716 wi83f06.x
	17	636.8	27.4	658	9	AV652040	AV652040 AV652040
	18	635.4	27.4	945	10	BG167885	BG167885 602339857
	19	625.6	27.0	724	10	BF031444	BF031444 601559244
	20	621.8	26.8	688	10	BG286763	BG286763 602382611
	21	618.8	26.7	796	10	BG432363	BG432363 602496688
c	22	604.8	26.1	861	9	AI829268	AI829268 wk58a10.x
	23	601	25.9	647	9	AV692287	AV692287 AV692287
	24	596	25.7	763	10	BG399019	BG399019 602440878
	25	594.2	25.6	663	9	AV646537	AV646537 AV646537
	26	592.4	25.5	859	10	BE249993	BE249993 600943171
c	27	589.8	25.4	640	9	AI635995	AI635995 wa14e12.x
c	28	588	25.3	683	9	AI480360	AI480360 tm51d11.x
	29	584.8	25.2	812	10	BG401258	BG401258 602465359
c	30	584	25.2	621	9	AW151709	AW151709 xf68b10.x
c	31	583	25.1	667	9	AW051715	AW051715 wx27g08.x
	32	583	25.1	969	10	BG291839	BG291839 602386068
c	33	572	24.7	636	9	AI767470	AI767470 wh27a05.x
	34	567.4	24.5	581	10	BE379923	BE379923 601159412
	35	567	24.4	623	10	BG171502	BG171502 602322042
	36	557.8	24.0	743	10	BG386668	BG386668 602454928
	37	556.4	24.0	741	10	BG433491	BG433491 602497921
c	38	550.6	23.7	619	9	AW300312	AW300312 xs59f05.x
	39	549.8	23.7	851	10	BG401340	BG401340 602465455
	40	543.8	23.4	700	10	BF130433	BF130433 601818721
c	41	542.8	23.4	608	9	AW469219	AW469219 hc79e05.x
	42	542.8	23.4	953	10	BG328092	BG328092 602427143
	43	537	23.1	935	10	BF234991	BF234991 602027544
c	44	535.4	23.1	626	9	AW082933	AW082933 xc04f05.x
	45	532.2	22.9	666	10	BF027719	BF027719 601764020

ALIGNMENTS

RESULT 1
BC011409
LOCUS BC011409 2450 bp mRNA linear HTC 30-JUL-2001
DEFINITION Homo sapiens, clone IMAGE:3854370, mRNA.
ACCESSION BC011409
VERSION BC011409.1 GI:15030273
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2450)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalon@bcm.tmc.edu.
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W., Muzny,D.M., Gibbs,R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 13 Row: b Column: 4
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4507814
This clone has the following problem: incomplete processing.
FEATURES Location/Qualifiers

```
Qy    1459 taccagtaccattccttggacgtgattggtttcctcttggccgtcgtgctgacagtggcc 1518
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 TACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGCC 1500

Qy    1519 ttcatcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcga 1578
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 TTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGGAAAAAAGGGCGA 1560

Qy    1579 gttaagaaagcccacaaatccaagacccattga 1611
           |||||||||||||||||||||||||||||||||
Db    1561 GTTAAGAAAGCCCACAAATCCAAGACCCATTGA 1593
```

Search completed: August 21, 2002, 11:58:40
Job time: 32903 sec

```
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1593)
  AUTHORS   Strassburg,C.P., Manns,M.P. and Tukey,R.H.
  TITLE     Expression of the UDP-glucuronosyltransferase 1A locus in human
            colon. Identification and characterization of the novel
            extrahepatic UGT1A8
  JOURNAL   J. Biol. Chem. 273 (15), 8719-8726 (1998)
  MEDLINE   98204857
   PUBMED   9535849
REFERENCE   2  (bases 1 to 1593)
  AUTHORS   Strassburg,C.P. and Tukey,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1997) Pharmacology and UCSD Cancer Center,
            University of California, San Diego, 9500 Gilman Dr, BSB 4th Floor
            Rm 4021, La Jolla, CA 92093-0636, USA
FEATURES             Location/Qualifiers
     source          1. .1593
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
     gene            1. .1593
                     /gene="UGT1A8"
     CDS             1. .1593
                     /gene="UGT1A8"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A8"
                     /protein_id="AAB84259.1"
                     /db_xref="GI:2613044"
                     /translation="MARTGWTSPIPLCVSLLLTCGFAEAGKLLVVPMDGSHWFTMQSV
                     VEKLILRGHEVVVVMPEVSWQLGKSLNCTVKTYSTSYTLEDLDREFMDFADAQWKAQV
                     RSLFSLFLSSSNGFFNLFFSHCRSLFNDRRLVEYLKESSFDAVFLDPFDACALIVAKY
                     FSLPSVVFARGIGCHYLEEGAQCPAPLSYVPRILLGFSDAMAFKERVRNHILHLEEHL
                     FCQYFSKNALEIASEILQTPVTAYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT      390 a    371 c    390 g    442 t
ORIGIN
```

Query Match 64.4%; Score 1493.8; DB 9; Length 1593;
Best Local Similarity 96.1%; Pred. No. 0;
Matches 1531; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```
Qy      19 atggcttgcacagggtggaccagccccttcctctatgtgtgtgtctgctgctgacctgt 78
           |||||| |||||||||||||||||||| |||| |||||||| | ||||||||||||||||
Db       1 ATGGCTCGCACAGGGTGGACCAGCCCCATTCCCCTATGTGTTTCTCTGCTGCTGACCTGT 60

Qy      79 ggctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcacc 138
           |||| ||| ||||||||||||| |||||||||||||||||||||||| ||||||||||||
Db      61 GGCTTTGCTGAGGCAGGGAAGCTGCTGGTAGTGCCCATGGATGGGAGTCACTGGTTCACC 120

Qy     139 atgaggtcggtggtggagaaactcattctcaggggggcatgaggtggttgtagtcatgcca 198
           |||  ||||||||||||||||| || ||||||||||||||||||||||||||||||||||
Db     121 ATGCAGTCGGTGGTGGAGAAACTTATCCTCAGGGGGCATGAGGTGGTTGTAGTCATGCCA 180

Qy     199 gaggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttca 258
           |||| |||||||||||||||| |||||||||||||||||||||||||||| ||||| |||
Db     181 GAGGTGAGTTGGCAACTGGGAAAATCACTGAATTGCACAGTGAAGACTTACTCAACCTCA 240

Qy     259 tataccctggaggatctggaccgggagttcaaggcttttgcccatgctcaatggaaagca 318
           || || ||||||||||||||||||||| |||| || ||| ||| ||||||||||||||||
Db     241 TACACTCTGGAGGATCTGGACCGGGAATTCATGGATTTCGCCGATGCTCAATGGAAAGCA 300

Qy     319 caagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttattt 378
           |||| |||||||| | | |||||||||  || |||||| |||||   ||||| ||||||||
Db     301 CAAGTACGAAGTTTGTTTTCTCTATTTCTGAGTTCATCCAATGGTTTTTTTAACTTATTT 360
```

```
Qy     379 ttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagagt 438
           |||||  |||||||||||||||||||| |||  || ||||||||||||||||||||||||
Db     361 TTTTCGCATTGCAGGAGTTTGTTTAATGACCGAAGATTAGTAGAATACTTAAAGGAGAGT 420

Qy     439 tcttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaatat 498
           |||||||||| |||||||| |||||||||||  |||||  |||||||||||||||||||
Db     421 TCTTTTGATGCGGTGTTTCTTGATCCTTTTGATGCCTGTGCGTTAATTGTTGCCAAATAT 480

Qy     499 ttctccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaaggt 558
           |||||||||||||| ||||||||||||||||||||  ||||||||||||||||||||||
Db     481 TTCTCCCTCCCCTCTGTGGTCTTCGCCAGGGGAATAGGTTGCCACTATCTTGAAGAAGGT 540

Qy     559 gcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgcc 618
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTCTTAGGGTTCTCAGATGCC 600

Qy     619 atgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatgc 678
           ||| |||||||||||||||||||||||||||| ||||||||||||||||||||| |||
Db     601 ATGGCTTTCAAGGAGAGAGTACGGAACCACATCTTGCACTTGGAGGAACATTTATTTTGC 660

Qy     679 caccgtttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttacg 738
           ||   ||||| ||||||||||||||||||||||||||||||||||||||||||||| ||
Db     661 CAGTATTTTTCCAAAAATGCCCTAGAAATAGCCTCTGAAATTCTCCAAACACCTGTCACA 720

Qy     739 gagtatgatctctacagccacacatcaatttggttgttgcgaacggacttttgttttggac 798
           |  ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     721 GCATATGATCTCTACAGCCACACATCAATTTGGTTGTTGCGAACAGACTTTGTTTTGGAC 780

Qy     799 tatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggga 858
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     781 TATCCCAAACCCGTGATGCCCAATATGATCTTCATTGGTGGTATCAACTGCCATCAGGGA 840

Qy     859 aagccgttgcctatggaatttgaagcctacattaatgcttctggagaacatggaattgtg 918
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 AAGCCATTGCCTATGGAATTTGAAGCCTACATTAATGCTTCTGGAGAACATGGAATTGTG 900

Qy     919 gttttctctttgggatcaatggtctcagaaattccagagaagaaagctatggcaattgct 978
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATGGCAATTGCT 960

Qy     979 gatgctttgggcaaaatccctcagacagtcctgtggcggtacactggaacccgaccatcg 1038
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGGAACCCGACCATCG 1020

Qy    1039 aatcttgcgaacaacacgatacttgttaagtggctaccccaaaacgatctgcttggtcac 1098
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 AATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACGATCTGCTTGGTCAC 1080

Qy    1099 ccgatgaccccgtgcctttatcacccatgctggttcccatggtgtttatgaaagcatatgc 1158
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAGCATATGC 1140

Qy    1159 aatggcgttcccatggtgatgatgcccttgtttggtgatcagatggacaatgcaaagcgc 1218
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 AATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAATGCAAAGCGC 1200

Qy    1219 atggagactaagggagctggagtgaccctgaatgttctggaaatgacttctgaagattta 1278
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGAAGATTTA 1260

Qy    1279 gaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctc 1338
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 GAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAACATCATGCGCCTC 1320

Qy    1339 tccagccttcacaaggaccgcccggtggagccgctggacctggccgtgttctgggtggag 1398
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 TCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAG 1380

Qy    1399 tttgtgatgaggcacaagggcgcgccacacctgcgccccgcagcccacgacctcacctgg 1458
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCCACGACCTCACCTGG 1440
```

```
Db    299 AATACTACACAAGAAAAATCTATCCAGTGCCGTATGACCAAGAAGAGCTGAAGAACCGTT 358

Qy    287 tcaaggcttttgcccatgctcaatggaaagcacaagtacgaagtatatattctctattaa 346

Db    359 ACCAATCATTTGGAAACAATCACTTTGCTGAGCGATCATTCCTAACTGCTCCTCAGACAG 418

Qy    347 tgggttcatacaatgaca---tttttgacttatttttttcaaattgcaggagtttgttta 403

Db    419 AGTACAGGAATAACATGATTGTTATTGGCCTGTACTTCATCAACTGCCAGAGCCTCCTGC 478

Qy    404 aagacaaaaaattagtagaatacttaaaggagagttcttttgatgcagtgtttctcgatc 463

Db    479 AGGACAGGGACACCCTGAACTTCTTTAAGGAGAGCAAGTTTGATGCTCTTTTCACAGACC 538

Qy    464 cttttgataactgtggcttaattgttgccaaatatttctccctcccctccgtggtcttcg 523

Db    539 CAGCCTTACCCTGTGGGGTGATCCTGGCTGAGTATTTGGGCCTACCATCTGTGTACCTCT 598

Qy    524 ccaggggaatactttgccactatcttgaagaaggtgcacagtgccctgctcctctttcct 583

Db    599 TCAGGGGTTTTCCGTGTTCCCTGGAGCATACATTCAGCAGAAGCCCAGACCCTGTGTCCT 658

Qy    584 atgtccccagaattctcttagggttctcagatgccatgactttcaaggagagagtacgga 643

Db    659 ACATTCCCAGGTGCTACACAAAGTTTTCAGACCACATGACTTTTTCCCAACGAGTGGCCA 718

Qy    644 accacatcatgcacttggaggaacatttattatgccaccgttttttcaaaaatgccctag 703

Db    719 ACTTCCTTGTTAATTTGTTGGAGCCCTATCTATTTTATTGTCTGTTTTCAAAGTATGAGA 778

Qy    704 aaatagcctctgaaattctccaaacacctgttacggagtatgatctctacagccacacat 763

Db    779 AACTCGCATCAGCTGTCCTCAAGAGAGATGT---GGATATAATCACCTTATCAGAGGTCT 835

Qy    764 caatttggttgttgcgaacggactttgttttggactatcccaaacccgtgatgcccaaca 823

Db    836 CTGTTTGGCTGTTAAGATATGACTTTGTGCTTGAATATCCTAGGCCGGTCATGCCCAACA 895

Qy    824 tgatcttcattggtggtatcaactgccatcagggaaagccgttgcctatggaatttgaag 883

Db    896 TGGTCTTCATTGGAGGTATCAACTGTAAGAAGAGGAAAGACTTGTCTCAGGAATTTGAAG 955

Qy    884 cctacattaatgcttctggagaacatggaattgtggtttctctttgggatcaatggtct 943

Db    956 CCTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCT 1015

Qy    944 cagaaattccagagaagaaagctatggcaattgctgatgctttgggcaaaatccctcaga 1003

Db   1016 CAGAAATTCCAGAGAAGAAAGCTATGGCAATTGCTGATGCTTTGGGCAAAAACCCTCAGA 1075

Qy   1004 cagtcctgtggcggtacactggaacccgaccatcgaatcttgcgaacaacacgatacttg 1063

Db   1076 CAGTCCTGTGGCGGTACACTGGAACCCGACCATCGAATCTTGCGAACAACACGATACTTG 1135

Qy   1064 ttaagtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcctttatcaccc 1123

Db   1136 TTAAGTGGCTACCCCAAAACGATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCC 1195

Qy   1124 atgctggttcccatggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgc 1183

Db   1196 ATGCTGGTTCCCATGGTGTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGC 1255

Qy   1184 ccttgtttggtgatcagatggacaatgcaaagcgcatggagactaagggagctggagtga 1243

Db   1256 CCTTGTTTGGTGATCAGATGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGA 1315

Qy   1244 ccctgaatgttctggaaatgacttctgaagatttagaaaatgctctaaaagcagtcatca 1303

Db   1316 CCCTGAATGTTCTGGAAATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCA 1375

Qy   1304 atgacaaaagttacaaggagaacatcatgcgcctctccagccttcacaaggaccgcccgg 1363

Db   1376 ATGACAAAAGTTACAAGGAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGG 1435

Qy   1364 tggagccgctggacctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgc 1423

Db   1436 TGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGC 1495

Qy   1424 cacacctgcgccccgcagcccacgacctcacctggtaccagtaccattccttggacgtga 1483

Db   1496 CACACCTGCGCCCCGCAGCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGA 1555

Qy   1484 ttggtttcctcttggccgtcgtgctgacagtggccttcatcacctttaaatgttgtgctt 1543

Db   1556 TTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTCCTT 1615

Qy   1544 atggctaccggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaaga 1603

Db   1616 ATGGCTACCCGAAATGCTTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGA 1675

Qy   1604 cccattgagaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttcca 1663

Db   1676 CCCATTGAGAGGTGGGTGGGAAATAAGGTAAAATTTTGAACCA-TCCCTAGTCATTTCCA 1734

Qy   1664 aacttgaaaacagaatcagtgttaaattcattttattcttattaaggaaatactttgcat 1723

Db   1735 AACTTGAAAACAGAATCAGTGTTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCAT 1794

Qy   1724 aaattaatcagccccagagtg-ctttaaaaaattctcttaaataaaaatagactcgc 1782

Db   1795 AAATTAATCAGCCCCAGAGTGTCTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGC 1854

Qy   1783 tagtcagtaaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatga 1842

Db   1855 TAGTCAGTAAAGATATTTGAATATGTATCGTGCCCCCTCCGGAGTCTTTGATCAGGATGA 1914

Qy   1843 catgtgccatttttcagaggacgtgcagacaggctggcattctagattactttttcttact 1902

Db   1915 CATGTGCCATTTTTCAGAGGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACT 1974

Qy   1903 ctgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctact 1962

Db   1975 CTGAAACATGGCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCACGGCTGCCCCTACT 2034

Qy   1963 gcaaatggcagttttaatcttatcttttggcttctgcagatggttgcaattgatccttaa 2022

Db   2035 GCAAATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTTGC-ATTGATCCTTAA 2093

Qy   2023 ccaataatggtcagtcctcatctctgtcctg-cttcataggtgccaccttgtgtgtttaa 2081

Db   2094 CCAATAATGGTCAGTCCTCATCTCTGTCCTGACTTCATAGGTGCCACCTTGTG-GTTTAA 2152

Qy   2082 agaagggaagctttgtaccttagagtgtaggtgaaatgaatgaatggcttggagtgcac 2141

Db   2153 AGAAGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCAC 2212

Qy   2142 tgagaacagcatatgatttcttgctttggggaaaaagaatgatgctatgaaattggtggg 2201

Db   2213 TGAGAACAGCATATGATTTCTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGG 2272

Qy   2202 tggtgtatttgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaacc 2261

Db   2273 TGGTGTATTTGAGAAGATAATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACC 2332

Qy   2262 caaaatacagctatgaagtgctgggcaagtttac-ttttttctgatgtttcctacaact 2320

Db   2333 CAAAATACAGCTATGAAGTGCTGGGCAAGTTTACTTTTTTTTCTGATGTTTCCTACAACT 2392

RESULT  15
AF030310
LOCUS       AF030310               1593 bp    mRNA    linear   PRI 26-OCT-2000
DEFINITION  Homo sapiens UDP-glucuronosyltransferase 1A8 (UGT1A8) mRNA,
            complete cds.
ACCESSION   AF030310
VERSION     AF030310.1  GI:2613043
```

```
Db   1256 CCTTGTTTGGTGATCAGATGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGA 1315

Qy   1244 ccctgaatgttctggaaatgacttctgaagatttagaaaatgctctaaaagcagtcatca 1303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1316 CCCTGAATGTTCTGGAAATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCA 1375

Qy   1304 atgacaaaagttacaaggagaacatcatgcgcctctccagccttcacaaggaccgccccgg 1363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1376 ATGACAAAAGTTACAAGGAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGG 1435

Qy   1364 tggagccgctggacctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgc 1423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1436 TGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGC 1495

Qy   1424 cacacctgcgccccgcagcccacgacctcacctggtaccagtaccattccttggacgtga 1483
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1496 CACACCTGCGCCCCGCAGCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGA 1555

Qy   1484 ttggtttcctcttggccgtcgtgctgacagtggccttcatcacctttaaatgttgtgctt 1543
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1556 TTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTCCTT 1615

Qy   1544 atggctaccggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaaga 1603
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1616 ATGGCTACCCGAAATGCTTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGA 1675

Qy   1604 cccattgagaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttcca 1663
          |||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1676 CCCATTGAGAGGTGGGTGGGAAATAAGGTAAAATTTTGAACCA-TCCCTAGTCATTTCCA 1734

Qy   1664 aacttgaaaacagaatcagtgttaaattcattttattcttattaaggaaatactttgcat 1723
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1735 AACTTGAAAACAGAATCAGTGTTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCAT 1794

Qy   1724 aaattaatcagccccagagtg-ctttaaaaaattctcttaaataaaaataatagactcgc 1782
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1795 AAATTAATCAGCCCCAGAGTGTCTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGC 1854

Qy   1783 tagtcagtaaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatga 1842
          ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1855 TAGTCAGTAAAGATATTTGAATATGTATCGTGCCCCCTCCGGAGTCTTTGATCAGGATGA 1914

Qy   1843 catgtgccatttttcagaggacgtgcagacaggctggcattctagattactttcttact 1902
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1915 CATGTGCCATTTTTCAGAGGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACT 1974

Qy   1903 ctgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctact 1962
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   1975 CTGAAACATGGCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCACCGGCTGCCCCTACT 2034

Qy   1963 gcaaatggcagttttaatcttatcttttggcttctgcagatggttgcaattgatccttaa 2022
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   2035 GCAAATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTTGC-ATTGATCCTTAA 2093

Qy   2023 ccaataatggtcagtcctcatctctgtcctg-cttcataggtgccaccttgtgtgtttaa 2081
          |||||||||||||||||||||||||||||| |||||||||||||||||||| ||||||
Db   2094 CCAATAATGGTCAGTCCTCATCTCTGTCCTGACTTCATAGGTGCCACCTTGTG-GTTTAA 2152

Qy   2082 agaagggaagctttgtacctttagagtgtaggtgaaatgaatgaatggcttggagtgcac 2141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2153 AGAAGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCAC 2212

Qy   2142 tgagaacagcatatgatttcttgctttggggaaaaagaatgatgctatgaaattggtggg 2201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2213 TGAGAACAGCATATGATTTCTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGG 2272

Qy   2202 tggtgtatttgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaacc 2261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2273 TGGTGTATTTGAGAAGATAATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACC 2332

Qy   2262 caaaatacagctatgaagtgctgggcaagtttac-ttttttctgatgtttcctacaact 2320
          |||| |||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   2333 CAAAATACAGCTATGAAGTGCTGGGCAAGTTTACTTTTTTTTCTGATGTTTCCTACAACT 2392
```

```
RESULT  14
HUMUGT1FA
LOCUS       HUMUGT1FA               2422 bp    mRNA    linear   PRI 14-JAN-1995
DEFINITION  Homo sapiens phenol UDP-glucuronosyltransferase (UDPGT) mRNA,
            complete cds.
ACCESSION   J04093
VERSION     J04093.1  GI:624724
KEYWORDS    UDP-glucuronosyltransferase; phenol.
SOURCE      Homo sapiens kidney cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2422)
  AUTHORS   Harding,D., Fournel-Gigleux,S., Jackson,M.R. and Burchell,B.
  TITLE     Cloning and substrate specificity of a human phenol
            UDP-glucuronosyltransferase expressed in COS-7 cells
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85 (22), 8381-8385 (1988)
  MEDLINE   89042187
COMMENT     On Jan 18, 1995 this sequence version replaced gi:340517.
FEATURES             Location/Qualifiers
     source          1. .2422
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="kidney"
     gene            1. .2422
                     /gene="UDPGT"
     CDS             88. .1683
                     /gene="UDPGT"
                     /EC_number="2.4.1.17"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase"
                     /protein_id="AAA61251.1"
                     /db_xref="GI:624725"
                     /translation="MACLLRSFQRISAGVFFLALWGMVVGDKLLVVPQDGSHWLSMKD
                     IVEVLSDRGHEIVVVVPEVNLLLKEYKYYTRKIYPVPYDQEELKNRYQSFGNNHFAER
                     SFLTAPQTEYRNNMIVIGLYFINCQSLLQDRDTLNFFKESKFDALFTDPALPCGVILA
                     EYLGLPSVYLFRGFPCSLEHTFSRSPDPVSYIPRCYTKFSDHMTFSQRVANFLVNLLE
                     PYLFYCLFSKYEKLASAVLKRDVDIITLSEVSVWLLRYDFVLEYPRPVMPNMVFIGGI
                     NCKKRKDLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKNPQTVLWR
                     YTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLF
                     GDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPV
                     EPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCC
                     PYGYPKCLGKKGRVKKAHKSKTH"
     polyA_signal    2396. .2401
                     /gene="UDPGT"
     polyA_signal    2404. .2409
                     /gene="UDPGT"
     polyA_site      2422
                     /gene="UDPGT"
BASE COUNT      653 a    516 c    555 g    698 t
ORIGIN
```

```
  Query Match           66.3%;  Score 1539.2;  DB 9;  Length 2422;
  Best Local Similarity  82.0%;  Pred. No. 0;
  Matches 1870;  Conservative    0;  Mismatches  398;  Indels   12;  Gaps    8;

Qy     47 ttcctctatgtgtgtctgctgacctgtggctttgccgaggcagggaagctactgg 106
          || ||  | | || |   |  |       || ||| | |  |  |  |  ||||| ||||
Db    119 TTTCTGCAGGGGTTTTCTTCTTAGCTCTTTGGGGCATGGTTGTAGGTGACAAGCTGCTGG 178

Qy    107 tagtgcccatggatgggagccactggttcaccatgaggtcggtggtggagaaactcattc 166
          | || ||   ||| || |||||||||| | |  |||| |    | || |||   |||| |
Db    179 TGGTCCCTCAGGACGGAAGCCACTGGCTTAGTATGAAGGATATAGTTGAGGTTCTCAGTG 238

Qy    167 tcaggggggcatgaggtggttgtagtcatgccagaggtgagttggcaactgggaagatcac 226
          | |||| |||||| | || || ||  |||| || || | || ||   ||  |  ||
Db    239 ACCGGGGTCATGAGATTGTAGTGGTGGTGCCTGAAGTTAATTTGCTTTTGAAAGAATACA 298

Qy    227 tgaattgcacagtgaagacttattcaacttcatataccctggaggatctggaccgggagt 286
            | | ||||   || |  ||| ||    | |||  ||  || || ||| |       |
```

```
Qy   1805 atgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggac 1864
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1811 ATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGAC 1870

Qy   1865 gtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttggga 1924
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1871 GTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGA 1930

Qy   1925 gtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatctta 1984
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1931 GTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTA 1990

Qy   1985 tcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatc 2044
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1991 TCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCTCATC 2050

Qy   2045 tctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtaccttta 2104
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2051 TCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTTA 2110

Qy   2105 gagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttcttg 2164
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2111 GAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTTCTTG 2170

Qy   2165 ctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatca 2224
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2171 CTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATAATCA 2230

Qy   2225 ttgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgctg 2284
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2231 TTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGTGCTG 2290

Qy   2285 ggcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||||
Db   2291 GGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 2326


RESULT  13
AX336711
LOCUS       AX336711                2422 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 7220 from Patent WO0194629.
ACCESSION   AX336711
VERSION     AX336711.1  GI:18127430
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 7220 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1. .2422
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      653 a    516 c    555 g    698 t
ORIGIN


  Query Match             66.3%;  Score 1539.2;  DB 6;  Length 2422;
  Best Local Similarity   82.0%;  Pred. No. 0;
  Matches 1870;  Conservative    0;  Mismatches  398;  Indels   12;  Gaps    8;

Qy      47 ttcctctatgtgtgtgtctgctgctgacctgtggctttgccgaggcagggaagctactgg 106
           ||  ||  | | ||  |   |  |       || ||| | |  |  |  |  |||| ||||
Db     119 TTTCTGCAGGGGTTTTCTTCTTAGCTCTTTGGGGCATGGTTGTAGGTGACAAGCTGCTGG 178

Qy     107 tagtgcccatggatgggagccactggttcaccatgaggtcggtggtggagaaactcattc 166
```

```
           | ||  ||   ||| || |||||||||.| |  |||| |    | ||  |||   |||| |
Db     179 TGGTCCCTCAGGACGGAAGCCACTGGCTTAGTATGAAGGATATAGTTGAGGTTCTCAGTG 238

Qy     167 tcaggggcatgaggtggttgtagtcatgccagaggtgagttggcaactgggaagatcac 226
           | |||| ||||||  | ||  ||  ||  |||| ||  ||  |  ||  ||   ||  | ||
Db     239 ACCGGGGTCATGAGATTGTAGTGGTGGTGCCTGAAGTTAATTTGCTTTTGAAAGAATACA 298

Qy     227 tgaattgcacagtgaagacttattcaacttcatataccctggaggatctggaccgggagt 286
             | | ||||   ||  |  |||  ||    | |||  ||  ||  ||  |||  |       |
Db     299 AATACTACACAAGAAAAATCTATCCAGTGCCGTATGACCAAGAAGAGCTGAAGAACCGTT 358

Qy     287 tcaaggcttttgcccatgctcaatggaaagcacaagtacgaagtatatattctctattaa 346
           | |  | ||||   |   ||| |     |  | |  |      |    | |||    |
Db     359 ACCAATCATTTGGAAACAATCACTTTGCTGAGCGATCATTCCTAACTGCTCCTCAGACAG 418

Qy     347 tgggttcatacaatgaca---tttttgacttatttttttcaaattgcaggagtttgttta 403
           |       | ||    |   || ||| | |  ||    || |||  |||  |  |
Db     419 AGTACAGGAATAACATGATTGTTATTGGCCTGTACTTCATCAACTGCCAGAGCCTCCTGC 478

Qy     404 aagacaaaaaattagtagaatacttaaaggagagttcttttgatgcagtgtttctcgatc 463
           | ||||   |     |  | | ||| ||||||||    ||||||||  | ||    || |
Db     479 AGGACAGGGACACCCTGAACTTCTTTAAGGAGAGCAAGTTTGATGCTCTTTTCACAGACC 538

Qy     464 cttttgataactgtggcttaattgttgccaaatatttctccctcccctccgtggtcttcg 523
           |        ||||||  | ||  | ||  | |||||  ||| || || |||  | ||
Db     539 CAGCCTTACCCTGTGGGGTGATCCTGGCTGAGTATTTGGGCCTACCATCTGTGTACCTCT 598

Qy     524 ccaggggaatactttgccactatcttgaagaaggtgcacagtgccctgctcctctttcct 583
           ||||||  | |  ||   |      |   |          |||| |  ||| | ||||
Db     599 TCAGGGGTTTTCCGTGTTCCCTGGAGCATACATTCAGCAGAAGCCCAGACCCTGTGTCCT 658

Qy     584 atgtccccagaattctcttagggttctcagatgccatgactttcaaggagagagtacgga 643
           |  | |||||      |  | ||| |||||  |||||||||     |  ||||   |
Db     659 ACATTCCCAGGTGCTACACAAAGTTTTCAGACCACATGACTTTTTCCCAACGAGTGGCCA 718

Qy     644 accacatcatgcacttggaggaacatttattatgccaccgttttttcaaaaatgccctag 703
           ||  | |  |  | |||  ||| |  |   |||   |  || | ||   |||
Db     719 ACTTCCTTGTTAATTTGTTGGAGCCCTATCTATTTTATTGTCTGTTTTCAAAGTATGAGA 778

Qy     704 aaatagcctctgaaattctccaaacacctgttacggagtatgatctctacagccacacat 763
           || | || || |   | ||| | | |  |||  |||          ||      |   |
Db     779 AACTCGCATCAGCTGTCCTCAAGAGAGATGT---GGATATAATCACCTTATCAGAGGTCT 835

Qy     764 caatttggttgttgcgaacggactttgttttggactatcccaaacccgtgatgcccaaca 823
           |  |||| ||||  ||  |||||||| | || ||||| |  || || ||||||||||
Db     836 CTGTTTGGCTGTTAAGATATGACTTTGTGCTTGAATATCCTAGGCCGGTCATGCCCAACA 895

Qy     824 tgatcttcattggtggtatcaactgccatcagggaaagccgttgcctatggaatttgaag 883
           || |||||||||| ||||||||||  |  || | ||   ||| ||  ||||||||||
Db     896 TGGTCTTCATTGGAGGTATCAACTGTAAGAAGAGGAAAGACTTGTCTCAGGAATTTGAAG 955

Qy     884 cctacattaatgcttctggagaacatggaattgtggttttctctttgggatcaatggtct 943
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     956 CCTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCT 1015

Qy     944 cagaaattccagagaagaaagctatggcaattgctgatgctttgggcaaaatccctcaga 1003
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    1016 CAGAAATTCCAGAGAAGAAAGCTATGGCAATTGCTGATGCTTTGGGCAAAAACCCTCAGA 1075

Qy    1004 cagtcctgtggcggtacactggaacccgaccatcgaatcttgcgaacaacacgatacttg 1063
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1076 CAGTCCTGTGGCGGTACACTGGAACCCGACCATCGAATCTTGCGAACAACACGATACTTG 1135

Qy    1064 ttaagtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcctttatcaccc 1123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1136 TTAAGTGGCTACCCCAAAACGATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCC 1195

Qy    1124 atgctggttcccatggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgc 1183
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1196 ATGCTGGTTCCCATGGTGTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGC 1255

Qy    1184 ccttgtttggtgatcagatggacaatgcaaagcgcatggagactaagggagctggagtga 1243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
                     /protein_id="AAA63195.1"
                     /db_xref="GI:184473"
                     /db_xref="GDB:G00-120-007"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQLQQRGHEIVVLAPDASLYIRDGAFYTLKTYPVPFQREDVKESFVSLGHNVFEN
                     DSFLQRVIKTYKKIKKDSAMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCSPIV
                     AQYLSLPTVFFLHALPCSLEFEATQCPNPFSYVPRPLSSHSDHMTFLQRVKNMLIAFS
                     QNFLCDVVYSPYATLASEFLQREVTVQDLLSSASVWLFRSDFVKDYPRPIMPNMVFVG
                     GINCLHQNPLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVL
                     WRYTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMP
                     LFGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDR
                     PVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFK
                     CCAYGYRKCLGKKGRVKKAHKSKTH"
     sig_peptide     43. .75
                     /gene="UGT1"
                     /note="G00-120-007"
     mat_peptide     76. .1614
                     /gene="UGT1"
                     /standard_name="bilirubin UDP-glucuronosyltransferase
                     isozyme 1"
                     /EC_number="2.4.1.17"
                     /note="G00-120-007"
                     /product="UDP-glucuronosyltransferase 1"
BASE COUNT      601 a    540 c    557 g    653 t
ORIGIN
```

Query Match 69.4%; Score 1610.8; DB 9; Length 2351;
Best Local Similarity 82.7%; Pred. No. 0;
Matches 1866; Conservative 0; Mismatches 387; Indels 3; Gaps 2;

```
Qy      68 tgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatgggagcc 127
Db      71 TGCTGGGCCCAGTGGTGTCCCATGCTGGGAAGATACTGTTGATCCCAGTGGATGGCAGCC 130

Qy     128 actggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgaggtggttg 187
Db     131 ACTGGCTGAGCATGCTTGGGGCCATCCAGCAGCTGCAGCAGAGGGGACATGAAATAGTTG 190

Qy     188 tagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaagactt 247
Db     191 TCCTAGCACCTGACGCCTCGTTGTACATCAGAGACGGAGCATTTTACACCTTGAAGACGT 250

Qy     248 attcaacttcatataccctggaggatctggaccgggagtt--caaggcttttgcccatgc 305
Db     251 ACCCTGTGCCATTCCAAAGGGAGGATGTGAAAGAGTCTTTTGTTAGTCTCGGGCATAATG 310

Qy     306 tcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgacat 365
Db     311 TTTTTGAGAATGATTCTTTCCTGCAGCGTGTGATCAAAACATACAAGAAAATAAAAAAGG 370

Qy     366 ttttgact-tatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaat 424
Db     371 ACTCTGCTATGCTTTTGTCTGGCTGTTCCCACTTACTGCACAACAAGGAGCTCATGGCCT 430

Qy     425 acttaaaggagagttctttttgatgcagtgtttctcgatccttttgataactgtggcttaa 484
Db     431 CCCTGGCAGAAAGCAGCTTTGATGTCATGCTGACGGACCCTTTCCTTCCTTGCAGCCCCA 490

Qy     485 ttgttgccaaaatatttctccctccccctccgtggtcttcgccaggggaatactttgccact 544
Db     491 TCGTGGCCCAGTACCTGTCTCTGCCCACTGTATTCTTCTTGCATGCACTGCCATGCAGCC 550

Qy     545 atcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttag 604
Db     551 TGGAATTTGAGGCTACCCAGTGCCCCAACCCATTCTCCTACGTGCCCAGGCCTCTCTCCT 610

Qy     605 ggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggagg 664
Db     611 CTCATTCAGATCACATGACCTTCCTGCAGCGGGTGAAGAACATGCTCATTGCCTTTTCAC 670

Qy     665 aacatttattatgccaccgtttttttcaaaaatgccctagaaatagcctctgaaattctcc 724
```

```
Db     671 AGAACTTTCTGTGCGACGTGGTTTATTCCCCGTATGCAACCCTTGCCTCAGAATTCCTTC 730

Qy     725 aaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacgg 784
Db     731 AGAGAGAGGTGACTGTCCAGGACCTATTGAGCTCTGCATCTGTCTGGCTGTTTAGAAGTG 790

Qy     785 actttgtttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatca 844
Db     791 ACTTTGTGAAGGATTACCCTAGGCCCATCATGCCCAATATGGTTTTTGTTGGTGGAATCA 850

Qy     845 actgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctggag 904
Db     851 ACTGCCTTCACCAAAATCCACTATCCCAGGAATTTGAAGCCTACATTAATGCTTCTGGAG 910

Qy     905 aacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaagaaag 964
Db     911 AACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAG 970

Qy     965 ctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactg 1024
Db     971 CTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTG 1030

Qy    1025 gaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacg 1084
Db    1031 GAACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACG 1090

Qy    1085 atctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgttt 1144
Db    1091 ATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTT 1150

Qy    1145 atgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatgg 1204
Db    1151 ATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGG 1210

Qy    1205 acaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatga 1264
Db    1211 ACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGA 1270

Qy    1265 cttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggaga 1324
Db    1271 CTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGA 1330

Qy    1325 acatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccg 1384
Db    1331 ACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCG 1390

Qy    1385 tgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagccc 1444
Db    1391 TGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCC 1450

Qy    1445 acgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcg 1504
Db    1451 ACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCG 1510

Qy    1505 tgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttgg 1564
Db    1511 TGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGG 1570

Qy    1565 ggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtggga 1624
Db    1571 GGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGA 1630

Qy    1625 aataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtg 1684
Db    1631 AATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTG 1690

Qy    1685 ttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtg 1744
Db    1691 TTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTG 1750

Qy    1745 ctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaat 1804
Db    1751 CTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAAT 1810
```

```
Db    726 AACTCGCATCAGCTGTCCTCAAGAGAGATGTGGATATAATCACCTTATATCAGAAGGTCT 785

Qy    764 caatttggttgttgcgaacggactttgttttggactatcccaaacccgtgatgcccaaca 823

Db    786 CTGTTTGGCTGTTAAGATATGACTTTGTGCTTGAATATCCTAGGCCGGTCATGCCCAACA 845

Qy    824 tgatcttcattggtggtatcaactgccatcagggaaagccgttgcctatggaatttgaag 883

Db    846 TGGTCTTCATTGGAGGTATCAACTGTAAGAAGAGGAAAGACTTGTCTCAGGAATTTGAAG 905

Qy    884 cctacattaatgcttctggagaacatggaattgtggttttctctttgggatcaatggtct 943

Db    906 CCTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCT 965

Qy    944 cagaaattccagagaagaaagctatggcaattgctgatgctttgggcaaaatccctcaga 1003

Db    966 CAGAAATTCCAGAGAAGAAAGCTATGGCAACTGCTGATGCTTTGGGCAAAATCCCTCAGA 1025

Qy   1004 cagtcctgtggcggtacactggaacccgaccatcgaatcttgcgaacaacacgatacttg 1063

Db   1026 CAGTCCTGTGGCGGTACACTGGAACCCGACCATCGAATCTTGCGAACAACACGATACTTG 1085

Qy   1064 ttaagtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcctttatcaccc 1123

Db   1086 TTAAGTGGCTACCCCAAAACGATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCC 1145

Qy   1124 atgctggttcccatggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgc 1183

Db   1146 ATGCTGGTTCCCATGGTGTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGC 1205

Qy   1184 ccttgtttggtgatcagatggacaatgcaaagcgcatggagactaagggagctggagtga 1243

Db   1206 CCTTGTTTGGTGATCAGATGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGA 1265

Qy   1244 ccctgaatgttctggaaatgacttctgaagatttagaaaatgctctaaaagcagtcatca 1303

Db   1266 CCCTGAATGTTCTGGAAATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCA 1325

Qy   1304 atgacaaaagttacaaggagaacatcatgcgcctctccagccttcacaaggaccgcccgg 1363

Db   1326 ATGACAAAAGTTACAAGGAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGG 1385

Qy   1364 tggagccgctggacctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgc 1423

Db   1386 TGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGC 1445

Qy   1424 cacacctgcgccccgcagcccacgacctcacctggtaccagtaccattccttggacgtga 1483

Db   1446 CACACCTGCGCCCCGCAGCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGA 1505

Qy   1484 ttggtttcctcttggccgtcgtgctgacagtggccttcatcacctttaaatgttgtgctt 1543

Db   1506 TTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTT 1565

Qy   1544 atggctaccggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaaga 1603

Db   1566 ATGGCTACCGGAAATGCTTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGA 1625

Qy   1604 cccattgagaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttcca 1663

Db   1626 CCCATTGAGAAGTGGGTGGGAAATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCA 1685

Qy   1664 aacttgaaaacagaatcagtgttaaattcattttattcttattaaggaaatactttgcat 1723

Db   1686 AACTTGAAAACAGAATCAGTGTTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCAT 1745

Qy   1724 aaattaatcagccccagagtgctttaaaaaattctcttaaataaaaataatagactcgct 1783

Db   1746 AAATTAATCAGCCCCAGAGTGCTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCT 1805

Qy   1784 agtcagtaaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgac 1843
```

```
Db   1806 AGTCAGTAAAGATATTTGAATATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGAC 1865

Qy   1844 atgtgccatttttcagaggacgtgcagacaggctggcattctagattacttttcttactc 1903

Db   1866 ATGTGCCATTTTTCAGAGGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTC 1925

Qy   1904 tgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctactg 1963

Db   1926 TGAAACATGGCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTG 1985

Qy   1964 caaatggcagttttaatcttatcttttggcttctgcagatggttgcaattgatccttaac 2023

Db   1986 CAAATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAAC 2045

Qy   2024 caataatggtcagtcctcatctctgtcctgcttcataggtgccaccttgtgtgtttaaag 2083

Db   2046 CAATAATGGTCAGTCCTCATCTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAG 2105

Qy   2084 aagggaagctttgtacctttagagtgtaggtgaaatgaatgaatggcttggagtgcactg 2143

Db   2106 AAGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTG 2165

Qy   2144 agaacagcatatgatttcttgctttggggaaaaagaatgatgctatgaaattggtgggtg 2203

Db   2166 AGAACAGCATATGATTTCTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTG 2225

Qy   2204 gtgtatttgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaaccca 2263

Db   2226 GTGTATTTGAGAAGATAATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCA 2285

Qy   2264 aaatacagctatgaagtgctgggcaagtttacttttttttctgatgtttcctacaact 2320

Db   2286 AAATACAGCTATGAAGTGCTGGGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 2342
```

```
RESULT  12
HUMHUGBR1
LOCUS       HUMHUGBR1               2351 bp    mRNA    linear   PRI 07-MAR-1995
DEFINITION  Human bilirubin UDP-glucuronosyltransferase isozyme 1 mRNA,
            complete cds.
ACCESSION   M57899
VERSION     M57899.1  GI:184472
KEYWORDS    UDP-glucuronosyltransferase; bilirubin UDP-glucuronosyltransferase.
SOURCE      Human adult female liver, cDNA to mRNA, clones Z6, Z11 and Z6MB2.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2351)
  AUTHORS   Ritter,J.K., Crawford,J.M. and Owens,I.S.
  TITLE     Cloning of two human liver bilirubin UDP-glucuronosyltransferase
            cDNAs with expression in COS-1 cells
  JOURNAL   J. Biol. Chem. 266 (2), 1043-1047 (1991)
  MEDLINE   91093210
FEATURES             Location/Qualifiers
     source          1. .2351
                     /organism="Homo sapiens"
                     /isolate="AK"
                     /db_xref="taxon:9606"
                     /map="2"
                     /clone="Z6"
                     /sex="female"
                     /tissue_type="liver"
                     /dev_stage="adult"
                     /tissue_lib="lambda-ZAP"
     gene            1. .2351
                     /gene="UGT1"
     CDS             16. .1617
                     /gene="UGT1"
                     /standard_name="bilirubin UDP-glucuronosyltransferase
                     isozyme 1"
                     /EC_number="2.4.1.17"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1"
```

```
Db   1901 GGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGAGTGCGGGATTCAA 1960

Qy   1938 aggtggtcccaccgctgcccctactgcaaatggcagttttaatcttatcttttggcttct 1997
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1961 AGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTATCTTTTGGCTTCT 2020

Qy   1998 gcagatggttgcaattgatccttaaccaataatggtcagtcctcatctctgtcctgcttc 2057
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2021 GCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCTCATCTCTGTCCTGCTTC 2080

Qy   2058 ataggtgccaccttgtgtgtttaaagaagggaagctttgtacctttagagtgtaggtgaa 2117
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2081 ATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAA 2140

Qy   2118 atgaatgaatggcttggagtgcactgagaacagcatatgatttcttgctttggggaaaaa 2177
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2141 ATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTTCTTGCTTTGGGGAAAAA 2200

Qy   2178 gaatgatgctatgaaattggtgggtggtgtatttgagaagataatcattgcttatgtcaa 2237
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2201 GAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATAATCATTGCTTATGTCAA 2260

Qy   2238 atggagctgaatttgataaaaacccaaaatacagctatgaagtgctgggcaagtttactt 2297
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2261 ATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGTGCTGGGCAAGTTTACTT 2320

Qy   2298 tttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||
Db   2321 TTTTTCTGATGTTTCCTACAACT 2343
```

```
RESULT  11
BC019861
LOCUS       BC019861               2388 bp    mRNA    linear   PRI 22-JAN-2002
DEFINITION  Homo sapiens, clone MGC:29860 IMAGE:4454322, mRNA, complete cds.
ACCESSION   BC019861
VERSION     BC019861.1  GI:18043700
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2388)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 42 Row: c Column: 8
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4507814.
FEATURES             Location/Qualifiers
     source          1. .2388
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:29860 IMAGE:4454322"
                     /tissue_type="Kidney, hypernephroma"
                     /clone_lib="NIH_MGC_89"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             35. .1633
                     /codon_start=1
                     /product="Unknown (protein for MGC:29860)"
                     /protein_id="AAH19861.1"
                     /db_xref="GI:18043701"
                     /translation="MACLLRSFQRISAGVFFLALWGMVVGDKLLVVPQDGSHWLSMKD
                     IVEVLSDRGHEIVVVVPEVNLLLKESKYYTRKIYPVPYDQEELKNRYQSFGNNHFAER
                     SFLTAPQTEYRNNMIVIGLYFINCQSLLQDRDTLNFFKESKFDALFTDPALPCGVILA
                     EYLGLPSVYLFRGFPCSLEHTFSRSPDPVSYIPRCYTKFSDHMTFSQRVANFLVNLLE
                     PYLFYCLFSKYEELASAVLKRDVDIITLYQKVSVWLLRYDFVLEYPRPVMPNMVFIGG
                     INCKKRKDLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMATADALGKIPQTVLW
                     RYTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPL
                     FGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRP
                     VEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKC
                     CAYGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT      648 a    503 c    546 g    691 t
ORIGIN
```

```
  Query Match             70.2%;  Score 1629;  DB 9;  Length 2388;
  Best Local Similarity   82.5%;  Pred. No. 0;
  Matches 1879;  Conservative    0;  Mismatches  395;  Indels    3;  Gaps    1;

Qy      47 ttcctctatgtgtgtctgctgctgacctgtggctttgccgaggcagggaagctactgg 106
           || ||  | | || |   |  |       || ||| | |  |  |  |  |||| ||||
Db      66 TTTCTGCAGGGGTTTTCTTCTTAGCACTTTGGGGCATGGTTGTAGGTGACAAGCTGCTGG 125

Qy     107 tagtgcccatggatgggagccactggttcaccatgaggtcggtggtggagaaactcattc 166
           | || ||   ||| || |||||||||| | |  |||| |    | || |||   |||| |
Db     126 TGGTCCCTCAGGACGGAAGCCACTGGCTTAGTATGAAGGATATAGTTGAGGTTCTCAGTG 185

Qy     167 tcaggggggcatgaggtggttgtagtcatgccagaggtgagttggcaactgggaagatcac 226
           | |||| |||||| | || || ||  |||| || || | | || ||   ||  |  |||
Db     186 ACCGGGGTCATGAGATTGTAGTGGTGGTGCCTGAAGTTAATTTGCTTTTGAAAGAATCCA 245

Qy     227 tgaattgcacagtgaagacttattcaacttcatataccctggaggatctggaccgggagt 286
             | | ||||   || |  ||| ||    | |||  ||  || || ||| |       |
Db     246 AATACTACACAAGAAAAATCTATCCAGTGCCGTATGACCAAGAAGAGCTGAAGAACCGTT 305

Qy     287 tcaaggctttgcccatgctcaatggaaagcacaagtacgaagtatatattctctattaa 346
           | |  | ||||   |  ||| |    | | |  |    |    |  |||   |
Db     306 ACCAATCATTTGGAAACAATCACTTTGCTGAGCGATCATTCCTAACTGCTCCTCAGACAG 365

Qy     347 tgggttcatacaatgaca---tttttgacttatttttttcaaattgcaggagtttgttta 403
           |       | ||    |   || ||| | | |  ||    || |||  |||  |  |
Db     366 AGTACAGGAATAACATGATTGTTATTGGCCTGTACTTCATCAACTGCCAGAGCCTCCTGC 425

Qy     404 aagacaaaaaattagtagaatacttaaaggagagttcttttgatgcagtgtttctcgatc 463
           | ||||   |    |  | | ||| ||||||||    ||||||||  | ||    || |
Db     426 AGGACAGGCACACCCTGAACTTCTTTAAGGAGAGCAAGTTTGATGCTCTTTTCACAGACC 485

Qy     464 cttttgataactgtggcttaattgttgccaaatatttctccctccccctccgtggtcttcg 523
           |         ||||||  | ||  | ||  | |||||   ||| || || |||  | ||
Db     486 CAGCCTTACCCTGTGGGGTGATCCTGGCTGAGTATTTGGGCCTACCATCTGTGTACCTCT 545

Qy     524 ccaggggaatactttgccactatcttgaagaaggtgcacagtgccctgctcctctttcct 583
           ||||||  | |  ||   |  |        |  |          |||| |  ||| | ||||
Db     546 TCAGGGGTTTTCCGTGTTCCCTGGAGCATACATTCAGCAGAAGCCCAGACCCTGTGTCCT 605

Qy     584 atgtccccagaattctcttagggttctcagatgccatgactttcaaggagagagtacgga 643
           |  | |||||      |  | ||| |||||   |||||||||     |  ||||   |
Db     606 ACATTCCCAGGTGCTACACAAAGTTTTCAGACCACATGACTTTTTCCCAACGAGTGGCCA 665

Qy     644 accacatcatgcacttggaggaacatttattatgccaccgtttttttcaaaaatgccctag 703
           ||  | |  |  | |||  ||| | |  |||   |  || | ||   |||      ||
Db     666 ACTTCCTTGTTAATTTGTTGGAGCCCTATCTATTTTATTGTCTGTTTTCAAAGTATGAAG 725

Qy     704 aaatagcctctgaaattctccaaacacctgttacggagtatgatctctacagccacacat 763
```

```
                     LAKYLSIPAVFFWRYIPCDLDFKGTQCPNPSSYIPKLLTTNSDHMTFLQRVKNMLYPL
                     ALSYICHTFSAPYASLASELFQREVSVVDLVSYASVWLFRGDFVMDYPRPIMPNMVFI
                     GGINCANGKPLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTV
                     LWRYTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMM
                     PLFGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKD
                     RPVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITF
                     KCCAYGYRKCLGKKGRVKKAHKSKTH"
     sig_peptide     63. .95
                     /gene="UGT2B"
                     /note="G00-127-753"
     mat_peptide     96. .1631
                     /gene="UGT2B"
                     /EC_number="2.4.1.17"
                     /note="G00-127-753"
                     /product="UDP-glucuronosyltransferase 2 family polypeptide
                     B"
BASE COUNT      604 a    534 c    573 g    657 t
ORIGIN
```

Query Match 70.5%; Score 1636.6; DB 9; Length 2368;
Best Local Similarity 83.4%; Pred. No. 0;
Matches 1871; Conservative 0; Mismatches 369; Indels 3; Gaps 1;

```
Qy      81 ctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcaccat 140
           ||  || |||    || ||| |  |||| |||||||  |||||  |||| |||| ||| |||
Db     101 CTGGGCTGAGAGTGGAAAGGTGTTGGTGGTGCCCACTGATGGCAGCCCCTGGCTCAGCAT 160

Qy     141 gaggtcggtggtggagaaactcattctcaggggcatgaggtggttgtagtcatgccaga 200
           | ||  ||   ||  | | |||  |  ||| || ||  ||| ||| ||  |||  |||||
Db     161 GCGGGAGGCCTTGCGGGAGCTCCATGCCAGAGGCCACCAGGCGGTGGTCCTCACCCCAGA 220

Qy     201 ggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttcata 260
           |||||  |  |||  |   |  |         || |||  |||   | ||| |   | |||
Db     221 GGTGAATATGCACATCAAAGAAGAGAAATTTTTCACCCTGACAGCCTATGCTGTTCCATG 280

Qy     261 taccctggaggatctggaccgggagttcaaggcttttgcccatgctcaat---ggaaagc 317
           ||||  |  ||||  |  ||  ||  |       ||  |   | ||  |     |    | ||
Db     281 GACCCAGAAGGAATTTGATCGCGTTACGCTGGGCTACACTCAAGGGTTCTTTGAAACAGA 340

Qy     318 acaagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttatt 377
           |||  |  |  |  |||||||  |   ||||        ||  |  | | |       |
Db     341 ACATCTTCTGAAGAGATATTCTAGAAGTATGGCAATTATGAACAATGTATCTTTGGCCCT 400

Qy     378 tttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagag 437
           |  |     |||  |    |  |  |  |  |     |  |     || | ||  |  |
Db     401 TCATAGGTGTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGGCACCTGAATGCTAC 460

Qy     438 ttcttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaata 497
           |||  |||||||  || ||    || ||  || |   |||  ||    |  | || || ||
Db     461 TTCCTTTGATGTGGTTTTAACAGACCCCGTTAACCTCTGCGGGGCGGTGCTGGCTAAGTA 520

Qy     498 tttctccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaagg 557
           |  ||  | ||  | |||  | ||    |||   || |  ||  |||         | ||
Db     521 CCTGTCGATTCCTGCTGTGTTTTTTTGGAGGTACATTCCATGTGACTTAGACTTTAAGGG 580

Qy     558 tgcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgc 617
             |||||||  ||   ||||   ||||||  | ||  |   | ||        |||||
Db     581 CACACAGTGTCCAAATCCTTCCTCCTATATTCCTAAGTTACTAACGACCAATTCAGACCA 640

Qy     618 catgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatg 677
           ||||||  |||  |  | ||  ||   ||||    ||    |   |||         |  | ||
Db     641 CATGACATTCCTGCAAAGGGTCAAGAACATGCTCTACCCTCTGGCCCTGTCCTACATTTG 700

Qy     678 ccaccgttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttac 737
           |||   |||||        |     |    | ||||||||  || | || | |   ||  |
Db     701 CCATACTTTTTCTGCCCCTTATGCAAGTCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTC 760

Qy     738 ggagtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttgga 797
           | |   |||||   |||| |  ||||  | ||| |||| ||||  ||||||||  ||||
Db     761 AGTGGTGGATCTTGTCAGCTATGCATCCGTGTGGCTGTTCCGAGGGGACTTTGTGATGGA 820

Qy     798 ctatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggg 857
```

```
           ||| ||||  ||  | |||||||||||| |||||||||| || ||||||||     | ||
Db     821 CTACCCCAGGCCGATCATGCCCAACATGGTCTTCATTGGGGGCATCAACTGTGCCAACGG 880

Qy     858 aaagccgttgcctatggaatttgaagcctacattaatgcttctggagaacatggaattgt 917
           |||||  |  ||  |||||||||||||||||||||||||||||||||||||||||||||
Db     881 GAAGCCACTATCTCAGGAATTTGAAGCCTACATTAATGCTTCTGGAGAACATGGAATTGT 940

Qy     918 ggttttctctttgggatcaatggtctcagaaattccagagaagaaaagctatggcaattgc 977
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     941 GGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATGGCAATTGC 1000

Qy     978 tgatgctttgggcaaaatccctcagacagtcctgtggcggtacactggaacccgaccatc 1037
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1001 TGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGGAACCCGACCATC 1060

Qy    1038 gaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacgatctgcttggtca 1097
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1061 GAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACGATCTGCTTGGTCA 1120

Qy    1098 cccgatgacccgtgcctttatcacccatgctggttcccatggtgtttatgaaagcatatg 1157
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1121 CCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAGCATATG 1180

Qy    1158 caatggcgttcccatggtgatgatgcccttgtttggtgatcagatggacaatgcaaagcg 1217
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1181 CAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAATGCAAAGCG 1240

Qy    1218 catggagactaagggagctggagtgaccctgaatgttctggaaatgacttctgaagattt 1277
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1241 CATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGAAGATTT 1300

Qy    1278 agaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcct 1337
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1301 AGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAACATCATGCGCCT 1360

Qy    1338 ctccagccttcacaaggaccgcccggtggagccgctggacctggccgtgttctgggtgga 1397
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1361 CTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGA 1420

Qy    1398 gtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagcccacgacctcacctg 1457
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1421 GTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCCACGACCTCACCTG 1480

Qy    1458 gtaccagtaccattccttggacgtgattggtttcctcttggccgtcgtgctgacagtggc 1517
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1481 GTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGC 1540

Qy    1518 cttcatcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcg 1577
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1541 CTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGGAAAAAAGGGCG 1600

Qy    1578 agttaagaaagcccacaaatccaagacccattgagaagtgggtgggaaataaggtaaaat 1637
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1601 AGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGAAATAAGGTAAAAT 1660

Qy    1638 tttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgttaaattcatttt 1697
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1661 TTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTGTTAAATTCATTTT 1720

Qy    1698 attcttattaaggaaatactttgcataaattaatcagccccagagtgctttaaaaaattc 1757
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1721 ATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTGCTTTAAAAAATTC 1780

Qy    1758 tcttaaataaaaaataatagactcgctagtcagtaaagatatttgaatatgtatcgtgccc 1817
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1781 TCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAATATGTATCGTGCCC 1840

Qy    1818 cctccggtgtctttgatcaggatgacatgtgccatttttcagaggacgtgcagacaggct 1877
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1841 CCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGACGTGCAGACAGGCT 1900

Qy    1878 ggcattctagattacttttcttactctgaaacatggcctgtttgggagtgcgggattcaa 1937
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy    846 ctgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctggaga 905
          ||| ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    481 CTGTCATCAGGGAAAGCCATTGCCTATGGAATTTGAAGCCTACATTAATGCTTCTGGAGA 540

Qy    906 acatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaagaaagc 965
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGC 600

Qy    966 tatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactgg 1025
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGG 660

Qy   1026 aacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacga 1085
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 AACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACGA 720

Qy   1086 tctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgttta 1145
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTA 780

Qy   1146 tgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatgga 1205
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGA 840

Qy   1206 caatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatgac 1265
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGAC 900

Qy   1266 ttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggagaa 1325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAA 960

Qy   1326 catcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccgt 1385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGT 1020

Qy   1386 gttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagccca 1445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCCA 1080

Qy   1446 cgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcgt 1505
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGT 1140

Qy   1506 gctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttggg 1565
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGG 1200

Qy   1566 gaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtgggaa 1625
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGAA 1260

Qy   1626 ataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgt 1685
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 ATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTGT 1320

Qy   1686 taaattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtgc 1745
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTGC 1380

Qy   1746 tttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaata 1805
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAATA 1440

Qy   1806 tgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggacg 1865
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TGTATCGTGCCCCCTCTGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGACG 1500

Qy   1866 tgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttgggag 1925
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGAG 1560

Qy   1926 tgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatcttat 1985
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TGCGGGATTCAAAGGTGGTCCCACGGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTAT 1620

Qy   1986 cttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatct 2045
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCTCATCT 1680

Qy   2046 ctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacctttag 2105
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CTGTCGTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTTAG 1740

Qy   2106 agtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttcttgc 2165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTTCTTGC 1800

Qy   2166 tttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatcat 2225
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATAATCAT 1860

Qy   2226 tgcttatgtgaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgctgg 2285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TGCTTATGTGAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGTGCTGG 1920

Qy   2286 gcaagtttactttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||
Db   1921 GCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 1955
```

```
RESULT  10
HUMHUGBR2
LOCUS       HUMHUGBR2              2368 bp    mRNA    linear   PRI 07-MAR-1995
DEFINITION  Human bilirubin UDP-glucuronosyltransferase isozyme 2 mRNA,
            complete cds.
ACCESSION   M57951
VERSION     M57951.1  GI:184474
KEYWORDS    UDP-glucuronosyltransferase; bilirubin UDP-glucuronosyltransferase.
SOURCE      Human adult female liver, cDNA to mRNA, clones Z17 and Z17P5E2.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2368)
  AUTHORS   Ritter,J.K., Crawford,J.M. and Owens,I.S.
  TITLE     Cloning of two human liver bilirubin UDP-glucuronosyltransferase
            cDNAs with expression in COS-1 cells
  JOURNAL   J. Biol. Chem. 266 (2), 1043-1047 (1991)
  MEDLINE   91093210
FEATURES             Location/Qualifiers
     source          1. .2368
                     /organism="Homo sapiens"
                     /isolate="AK"
                     /db_xref="taxon:9606"
                     /map="Unassigned"
                     /clone="Z17"
                     /sex="female"
                     /tissue_type="liver"
                     /dev_stage="adult"
                     /tissue_lib="lambda-ZAP"
     gene            1. .2368
                     /gene="UGT2B"
     CDS             30. .1634
                     /gene="UGT2B"
                     /standard_name="bilirubin UDP-glucuronosyltransferase
                     isozyme 2"
                     /EC_number="2.4.1.17"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 2 family polypeptide
                     B"
                     /protein_id="AAA63196.1"
                     /db_xref="GI:184475"
                     /db_xref="GDB:G00-127-753"
                     /translation="MARGLQVPLPRLATGLLLLLSVQPWAESGKVLVVPTDGSPWLSM
                     REALRELHARGHQAVVLTPEVNMHIKEEKFFTLTAYAVPWTQKEFDRVTLGYTQGFFE
                     TEHLLKRYSRSMAIMNNVSLALHRCCVELLHNEALIRHLNATSFDVVLTDPVNLCGAV
```

```
Qy   1502 tcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgct 1561
          |||||||||||||||||||||||| ||||||||||||||| ||||||||||||||||||||
Db   1543 TCGTGCTGACAGTGGCCTTCATCGCCTTTAAATGTTGTGCCTATGGCTACCGGAAATGCT 1602

Qy   1562 tggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtg 1621
          | |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   1603 TCGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTTGGTG 1662

Qy   1622 ggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatca 1681
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   1663 GGAAATAAGGTAAAATTTTGAACCATTCCCTAGTCAGTTCCAAACTTGAAAACAGAATCA 1722

Qy   1682 gtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccaga 1741
          |||||||||||||||||||   |||||||||||||||||||||||||||||||||||||
Db   1723 GTGTTAAATTCATTTTATTC---TTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGA 1779

Qy   1742 gtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttg 1801
          |||||||||||||||| ||||||||||||||||||| || ||||||||||||||||||||
Db   1780 GTGCTTTAAAAAATTATCTTAAATAAAAATAATAGAATCACTAGTCAGTAAAGATATTTG 1839

Qy   1802 aatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagag 1861
          |||||||||| || |||||||||||||||| ||||| ||||||| |||||| ||||||||
Db   1840 AATATGTATCATG-CCCCTCCGGTGTCTTCGATCAAGATGACACATGCCATCTTTCAGAG 1898

Qy   1862 gacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttg 1921
          || |||||||||||||||| ||||| ||  |||||||||||| |||||||| ||||||||
Db   1899 GATGTGCAGACAGGCTGGCGTTCTACATGCCTTTTCTTACTCGGAAACATGACCTGTTTG 1958

Qy   1922 ggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatc 1981
          |||||||| |||||||||||||||||||||||||| ||||| ||||||||||||||||||||
Db   1959 GGAGTGCAGGATTCAAAGGTGGTCCCACCGCTGTCCCTATTGCAAATGGCAGTTTTAATC 2018

Qy   1982 ttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctc 2041
          |||||||||||||||| |||||||||| |||||||||||||||||||||||  || |||
Db   2019 TTATCTTTTGGCTTCTGTAGATGGTTGCGATTGATCCTTAACCAATAATGGTTGGTTCTC 2078

Qy   2042 atctctgtcc-tgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacc 2100
          |||||||||| |||||||||| ||| |||||||||||||||||||||| ||||||||| 
Db   2079 ATCTCTGTCCGTGCTTCATAGGCGCCCCCTTGTGTGTTTAAAGAAGGGAAACTTTGTACC 2138

Qy   2101 tttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttt 2160
          |||||||| |||||||||||| |||||||||||||||||||||||||||||||| ||| |
Db   2139 TTTAGAGCATAGGTGAAATGAGTGAATGGCTTGGAGTGCACTGAGAACAGCATACGATCT 2198

Qy   2161 cttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagata 2220
          |||||||| |||||||||||||||||| | |||||| |||||||||| |||||||||||
Db   2199 ATTGCTTTGAGGAAAAAGAATGATGCTGTAAAATTGATGGGTGGTGTGTTTGAGAAGATA 2258

Qy   2221 atcattgcttatgtcaaatggagctgaatttgataaaaaacccaaaatacagctatgaagt 2280
          |||||||||||||||||||||||||||||||| ||||||||||||||||||  |||||||
Db   2259 ATCATTGCTTATGTCAAATGGAGCTGAATTTCATAAAAACCCAAAATACAGTGATGAAGT 2318

Qy   2281 gctgggcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||| || ||||||||| ||||||||| |||||
Db   2319 GCTGGGCAAGCGTA-TTTTTTTCTTATGTTTCCTGCAACT 2357
```

```
RESULT   9
AK025403
LOCUS       AK025403               2003 bp    mRNA    linear   PRI 29-SEP-2000
DEFINITION  Homo sapiens cDNA: FLJ21750 fis, clone COLF5999, highly similar to
            AF056188 Homo sapiens UDP-glucuronosyltransferase 1A9 (UGT1A9)
            mRNA.
ACCESSION   AK025403
VERSION     AK025403.1  GI:10437911
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens colon mucosa cDNA to mRNA, clone_lib:ColF
            clone:COLF5999.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
            Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
            Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 2003)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
            Sugano, Institute of Medical Science, University of Tokyo,
            Laboratory of Genome Structure Analysis, Human Genome Center;
            Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Departent of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1. .2003
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="COLF5999"
                     /clone_lib="ColF"
                     /tissue_type="colon mucosa"
                     /note="cloning vector pME18SFL3"
     misc_feature    1. .2003
                     /note="highly similar to AF056188 Homo sapiens
                     UDP-glucuronosyltransferase 1A9 (UGT1A9) mRNA"
BASE COUNT      551 a    414 c    445 g    593 t
ORIGIN
```

Query Match 81.7%; Score 1895.8; DB 9; Length 2003;
Best Local Similarity 98.1%; Pred. No. 0;
Matches 1918; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
Qy    366 ttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaata 425
          | ||||||||||||||||  |||||||||||||||||||| |||  ||||||||||||||||
Db      1 TCTTGACTTATTTTTTTCGCATTGCAGGAGTTTGTTTAATGACCGAAAATTAGTAGAATA 60

Qy    426 cttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggcttaat 485
          ||||||||||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||
Db     61 CTTAAAGGAGAGTTCTTTTGATGCAGTGTTTCTGGATCCTTTTGATACCTGTGGCTTAAT 120

Qy    486 tgttgccaaatatttctccctccccctccgtggtcttcgccaggggaatactttgccacta 545
          |||||| |||||||||||||||||| ||||||||| |||||||||| |||||||| |
Db    121 TGTTGCTAAATATTTCTCCCTCCCCTCTGTGGTCTTCACCAGGGGAATATTTTGCCACCA 180

Qy    546 tcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttagg 605
          ||||||||||||||||||||||||||||||||||||||||||||||||||    |||||||||
Db    181 TCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAATGATCTCTTAGG 240

Qy    606 gttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggagga 665
          |||||||||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||
Db    241 GTTCTCAGATGCCATGACTTTCAAGGAGAGAGTATGGAACCACATCGTGCACTTGGAGGA 300

Qy    666 acatttattatgccaccgtttttttcaaaaatgccctagaaatagcctctgaaattctcca 725
          |||||||| |||||   | |||| | ||||||||||||||||||||||||||||||||||||
Db    301 CCATTTATTTTGCCAGTATCTTTTTAGAAATGCCCTAGAAATAGCCTCTGAAATTCTCCA 360

Qy    726 aacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacgga 785
          ||| ||||| |||| ||||||||||||| ||||||||||||||||||||||||||||||||
Db    361 AACCCCTGTCACGGCATATGATCTCTACAGTCACACATCAATTTGGTTGTTGCGAACGGA 420

Qy    786 ctttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaa 845
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTTTGTTTTGGACTATCCCAAACCCGTGATGCCCAACATGATCTTCATTGGTGGTATCAA 480
```

```
VERSION     AF104336.1  GI:6537137
KEYWORDS    .
SOURCE      crab-eating macaque.
  ORGANISM  Macaca fascicularis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 2396)
  AUTHORS   Albert,C., Vallee,M., Beaudry,G., Belanger,A. and Hum,D.W.
  TITLE     The monkey and human uridine diphosphate-glucuronosyltransferase
            UGT1A9, expressed in steroid target tissues, are
            estrogen-conjugating enzymes
  JOURNAL   Endocrinology 140 (7), 3292-3302 (1999)
  MEDLINE   99312152
   PUBMED   10385426
REFERENCE   2  (bases 1 to 2396)
  AUTHORS   Albert,C., Vallee,M., Beaudry,G., Belanger,A. and Hum,D.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-1998) Molecular Endocrinology, Laval University,
            2705 Laurier Boulevard, Sainte-Foy, Que G1V4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .2396
                     /organism="Macaca fascicularis"
                     /db_xref="taxon:9541"
     CDS             60. .1652
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase UGT1A09"
                     /protein_id="AAF15546.1"
                     /db_xref="GI:6537138"
                     /translation="MTRTGWSSPLPLCVCLLLACGFAEAGKLLVVPVDGSHWLSMRSV
                     VEKLILRGHEVVVVMPKVSWQLGRPLNCTVKTYSTSYTLEDLDREFKAFRQAQWKAQV
                     QSIFSLIMSSYNDIFDLFFSNCRSLFKDKKLVEYLKESSFDAVFLDPFDTCGLIVAKY
                     LSLPSVVFARGVFCHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEERL
                     LCHRFFKSALEIASEILQTPVTVYDLHRQISIWLLRMDFVLDYPKPVMPNMVFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVAEIPEKKAMAIADALGKIPQTVLWRY
                     TGTPPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGIYEGICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMHLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAIVLTVAFIAFKCCA
                     YGYRKCFGKKGRVKKAHKSKTH"
BASE COUNT      645 a    513 c    558 g    680 t
ORIGIN
```

```
  Query Match           89.5%;  Score 2076;  DB 9;  Length 2396;
  Best Local Similarity  94.6%;  Pred. No. 0;
  Matches 2194;  Conservative    0;  Mismatches  120;  Indels    6;  Gaps    4;

Qy       2 tcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtgt 61
          || ||||| |||| |||||| || ||||||||||| ||| |||||||| ||||||||||
Db      43 TCGGCTGCGGTTCCCTGATGACTCGCACAGGGTGGAGCAGTCCCCTTCCCCTATGTGTGT 102

Qy      62 gtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatg 121
           |||||||||| |||||||||||||||||||||||||||||| |||||||||||| ||||||
Db     103 GTCTGCTGCTGGCCTGTGGCTTTGCCGAGGCAGGGAAGCTGCTGGTAGTGCCCGTGGATG 162

Qy     122 ggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgaagg 181
           |||||||||| || ||||||||||||||||||||||||||||||||||||||||||||||
Db     163 GGAGCCACTGGCTCTCCATGAGGTCGGTGGTGGAGAAACTCATTCTCAGGGGGCATGAGG 222

Qy     182 tggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtga 241
           |||| |||||||||||| |||||||||||| ||||||||| ||||||||||||||||||
Db     223 TGGTCGTAGTCATGCCAAAGGTGAGTTGGCAGCTGGGAAGACCACTGAATTGCACAGTGA 282

Qy     242 agacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgccc 301
           |||||||||||||||||| || ||||||||||||||||||||||||||||||||||  ||
Db     283 AGACTTATTCAACTTCATACACTCTGGAGGATCTGGACCGGGAGTTCAAGGCTTTTCGCC 342

Qy     302 atgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatg 361
           | |||||||||||||||||||| |||||||| ||||||| ||||| |||| |||||||
Db     343 AGGCTCAATGGAAAGCACAAGTACAAAGTATATTTTCTCTAATAATGAGTTCCTACAATG 402

Qy     362 acatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtag 421
           |||||||||||||||||||||||||||||||||||||||||||||||| |||||||| |||||||

Db     403 ACATTTTTGACTTATTTTTTTCAAATTGCAGGAGTTTGTTTAAGGACAAAAAGTTAGTAG 462

Qy     422 aatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggct 481
           |||||||||||||||||||||||||||||||| || ||||| |||||||||||| |||||||
Db     463 AATACTTAAAGGAGAGTTCTTTTGATGCGGTATTTCTGGATCCTTTTGATACCTGTGGCT 522

Qy     482 taattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgcc 541
           |||| |||||||||||| |||||||||||| |||||||||||||||||| || ||||||
Db     523 TAATCGTTGCCAAATATCTCTCCCTCCCCTCTGTGGTCTTCGCCAGGGGAGTATTTTGCC 582

Qy     542 actatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctct 601
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     583 ACTATCTTGAAGAAGGTGCGCAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTCT 642

Qy     602 tagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttgg 661
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     643 TAGGGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTACGGAACCACATCATGCACTTGG 702

Qy     662 aggaacatttattatgccaccgtttttttcaaaaatgccctagaaatagcctctgaaattc 721
           |||||| |||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     703 AGGAACGTTTATTATGCCACCGTTTTTTCAAAAGTGCCCTAGAAATAGCCTCTGAAATTC 762

Qy     722 tccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaa 781
           | |||||||||| || |  |||||||  |||| || | |||||||||||||| ||||||
Db     763 TTCAAACACCTGTCACAGTATATGATCTTCACAGACAAATATCAATTTGGTTGCTGCGAA 822

Qy     782 cggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggta 841
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     823 TGGACTTTGTTTTGGACTATCCCAAACCCGTGATGCCCAACATGGTCTTCATTGGTGGTA 882

Qy     842 tcaactgcccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctg 901
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     883 TCAACTGCCATCAGGGAAAGCCATTGCCTATGGAATTTGAAGCCTACATTAATGCTTCTG 942

Qy     902 gagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaaga 961
           |||||||| |||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     943 GAGAACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCGCAGAAATTCCAGAGAAGA 1002

Qy     962 aagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtaca 1021
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1003 AAGCTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACA 1062

Qy    1022 ctggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaa 1081
           |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1063 CTGGAACCCCACCATCGAATCTTGCGAACAATACGATACTTGTTAAGTGGCTACCCCAAA 1122

Qy    1082 acgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtg 1141
           ||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1123 ACGATCTTCTTGGTCACCCAATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTA 1182

Qy    1142 tttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcaga 1201
           |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1183 TTTATGAAGGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGA 1242

Qy    1202 tggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaa 1261
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    1243 TGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTTGAAA 1302

Qy    1262 tgacttctcaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaagg 1321
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    1303 TGACTTCTCAAGATTTAGAAAATGCCCTAAAAGCAGTCATCAATGACAAAAGCTACAAGG 1362

Qy    1322 agaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctgg 1381
           |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    1363 AGAACATCATGCACCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGG 1422

Qy    1382 ccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcag 1441
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    1423 CCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCACCACACCTGCGCCCCGCAG 1482

Qy    1442 cccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccg 1501
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||| |||||
Db    1483 CCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATCGGTTTCCTCCTGGCCA 1542
```

```
Db    265 ATTCAACTTCATATA-CCTGGAGGATCTGGACCGGGAGTTCAAGGCTTTTGCCCATGCTC 323
Qy    308 aatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgacattt 367
Db    324 AATGGAAAGCACAAGTACGAAGTATATATTCTCTATTAATGGGTTCATACAATGACATTT 383
Qy    368 ttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatact 427
Db    384 TTGACTTATTTTTTTCAAATTGCAGGAGTTTGTTTAAAGACAAAAAATTAGTAGAATACT 443
Qy    428 taaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggcttaattg 487
Db    444 TAAAGGAGAGTTCTTTTGATGCAGTGTTTCTCGATCCTTTTGATAACTGTGGCTTAATTG 503
Qy    488 ttgccaaatatttctccctccccctccgtggtcttcgccaggggaatactttgccactatc 547
Db    504 TTGCCAAATATTTCTCCCTCCCCTCCGTGGTCTTCGCCAGGGGAATACTTTGCCACTATC 563
Qy    548 ttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttagggt 607
Db    564 TTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTCTTAGGGT 623
Qy    608 tctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggaggaac 667
Db    624 TCTCAGATGACATGACTTTCAAGGAGAGAGTACGGAACCACATCATGCACTTGGAGGAAC 683
Qy    668 atttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaattctccaaa 727
Db    684 ATTTATTATGCCACCGTTTTTTCAAAAATGCCCTAGAAATAGCCTCTGAAATTCTCCAAA 743
Qy    728 cacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacggact 787
Db    744 CACCTGTTACGGAGTATGATCTCTACAGCCACACATCAATTTGGTTGTTGCGAACGGACT 803
Qy    788 ttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaact 847
Db    804 TTGTTTTGGACTATCCCAAACCCGTGATGCCCAACATGATCTTCATTGGTGGTATCAACT 863
Qy    848 gccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctggagaac 907
Db    864 GCCATGAGAGGAAAGCGTTGCCTATGGAATTTGAAGCCTACATTAATGCTTCTGGAGAAC 923
Qy    908 atggaattgtggttttctctttgggatcaatggtctcagaaattccagagaagaaagcta 967
Db    924 ATGGAATTGTGGGTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTA 983
Qy    968 tggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactggaa 1027
Db    984 TGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGGAA 1043
Qy   1028 cccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacgatc 1087
Db   1044 CCCGACCATGCAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACGATC 1103
Qy   1088 tgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgtttatg 1147
Db   1104 TGCTTGGTCACCCGATGACCAGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATG 1163
Qy   1148 aaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatggaca 1207
Db   1164 AAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACA 1223
Qy   1208 atgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatgactt 1267
Db   1224 ATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCATGAATGTTCTGGAAATGACTT 1283
Qy   1268 ctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggagaaca 1327
Db   1284 CTGAAGAATTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAACA 1343
Qy   1328 tcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccgtgt 1387
Db   1344 TCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGT 1403
Qy   1388 tctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagcccacg 1447
Db   1404 TCTGGGTGGAGTTTGTTATGAGGCACAAGGGCGCGACACACCTGCGCCCCGCAGCCCACG 1463
Qy   1448 acctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcgtgc 1507
Db   1464 ACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGCCGTGC 1523
Qy   1508 tgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttgggga 1567
Db   1524 TGACAGTGGCCTTCATCACCTGTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGGA 1583
Qy   1568 aaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtgggaaat 1627
Db   1584 AAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAGGTGGGTGGGAAAT 1643
Qy   1628 aaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgtta 1687
Db   1644 AAGGTAAAATTTTGAACCA-TCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTGTTA 1702
Qy   1688 aattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtg-ct 1746
Db   1703 AATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTGTCT 1762
Qy   1747 ttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaatat 1806
Db   1763 TTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAATAT 1822
Qy   1807 gtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggacgt 1866
Db   1823 GTATCGTGCCCCCTCCGGAGTCTTTGATCAGGATGACATGTGCCATTTTTCAAAGGACGT 1882
Qy   1867 gcagacaggctggcattc-tagattacttttcttactctgaaacatggcctgtttgggag 1925
Db   1883 GCAGACAGGCTGGCATTCTTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGAG 1942
Qy   1926 tgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatcttat 1985
Db   1943 TGCGGGATTCAAAGGTGGTCCCACGGCTGCCCCTACTGCAAATGGGAGTTTTAATCCTAT 2002
Qy   1986 cttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatct 2045
Db   2003 CTTTTGGCTTCTGCAGATGGTTGC-ATTGATCCTTAACCAATAATGGTCAGTCCTCATCT 2061
Qy   2046 ctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacctttag 2105
Db   2062 CTGTCCTGACTTCATAGGTGCCAGGTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTTAG 2121
Qy   2106 agtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttcttgc 2165
Db   2122 AGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTTCTTGC 2181
Qy   2166 tttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatcat 2225
Db   2182 TTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTTTATTTGAGAAGATAATCAT 2241
Qy   2226 tgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgctgg 2285
Db   2242 TGCTTATGTCAAATGGAGGTTAATTTGATAAAAACCCAAAATACAGCTATGAAGTGCTGG 2301
Qy   2286 gcaagtttac-ttttttttctgatgtttcctacaact 2320
Db   2302 GCAAGTTTACTTTTTTTTCTGATGTTTCCTACAAAT 2337
```

RESULT 8
AF104336
LOCUS AF104336 2396 bp mRNA linear PRI 13-DEC-1999
DEFINITION Macaca fascicularis UDP-glucuronosyltransferase UGT1A09 mRNA, complete cds.
ACCESSION AF104336

```
Db   1196 ATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGG 1255

Qy   1205 acaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatga 1264
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1256 ACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGA 1315

Qy   1265 cttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggaga 1324
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1316 CTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGA 1375

Qy   1325 acatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccg 1384
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1376 ACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCG 1435

Qy   1385 tgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagccc 1444
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1436 TGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCC 1495

Qy   1445 acgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcg 1504
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1496 ACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCG 1555

Qy   1505 tgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttgg 1564
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1556 TGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGG 1615

Qy   1565 ggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtggga 1624
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1616 GGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGA 1675

Qy   1625 aataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtg 1684
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1676 AATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTG 1735

Qy   1685 ttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtg 1744
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1736 TTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTG 1795

Qy   1745 cttt-aaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaa 1803
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1796 CTTTAAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAA 1855

Qy   1804 tatgcatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagagga 1863
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1856 TATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGA 1915

Qy   1864 cgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttggg 1923
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1916 CGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGG 1975

Qy   1924 agtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatctt 1983
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1976 AGTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAATCTT 2035

Qy   1984 atcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcat 2043
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2036 ATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCTCAT 2095

Qy   2044 ctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacctt 2103
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2096 CTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTT 2155

Qy   2104 agagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttctt 2163
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2156 AGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTTCTT 2215

Qy   2164 gctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatc 2223
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2216 GCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATAATC 2275

Qy   2224 attgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgct 2283
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2276 ATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGTGCT 2335

Qy   2284 gggcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||||
Db   2336 GGGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 2372
```

```
RESULT   7
S55985
LOCUS       S55985                2392 bp    mRNA    linear   PRI 04-MAY-2000
DEFINITION  Homo sapiens UDP-glucuronosyltransferase mRNA, complete cds.
ACCESSION   S55985
VERSION     S55985.1  GI:235453
KEYWORDS    .
SOURCE      human liver.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2392)
  AUTHORS   Wooster,R., Sutherland,L., Ebner,T., Clarke,D., Da Cruz e Silva,O.
            and Burchell,B.
  TITLE     Cloning and stable expression of a new member of the human liver
            phenol/bilirubin: UDP-glucuronosyltransferase cDNA family
  JOURNAL   Biochem. J. 278 (Pt 2), 465-469 (1991)
  MEDLINE   91378895
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 55985] from the original journal article.
FEATURES             Location/Qualifiers
     source          1. .2392
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
     CDS             35. .1627
                     /note="phenol/bilirubin; lugP4; conceptual translation
                     presented here differs from translation in publication"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase"
                     /protein_id="AAB19791.2"
                     /db_xref="GI:7690346"
                     /translation="MACTGWTSPLPLCVCLLLTCGFAEAGKLVVVPMDGSHWFTMRSV
                     VEKLILRGHEVVVVSARGELATGKINELHSEDLFNFIYLEDLDREFKAFAHAQWKAQV
                     RSIYSLLMGSYNDIFDLFFSNCRSLFKDKKLVEYLKESSFDAVFLDPFDNCGLIVAKY
                     FSLPSVVFARGILCHYLEEGAQCPAPLSYVPRILLGFSDDMTFKERVRNHIMHLEEHL
                     LCHRFFKNALEIASEILQTPVTEYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHERKALPMEFEAYINASGEHGIVGFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPCNLANNTILVKWLPQNDLLGHPMTSAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTMNVLEMTSEELENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGATHLRPAAHDLTWYQYHSLDVIGFLLAAVLTVAFITCKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT      661 a    492 c    554 g    685 t
ORIGIN
```

```
  Query Match             93.0%;  Score 2156.8;  DB 9;  Length 2392;
  Best Local Similarity   97.7%;  Pred. No. 0;
  Matches 2262;  Conservative    0;  Mismatches   47;  Indels    7;  Gaps    7;

Qy       9 cagttctctgatggcttgcacagggtggaccagccccctcctctatgtgtgtgtctgct 68
           ||  |||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25 CACCTCTGGCATGGCTTGCACAGGGTGGACCAGCCCCCTTCCTCTATGTGTGTGTCTGCT 84

Qy      69 gctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatgggagcca 128
           ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      85 GCTGACCTGTGGCTTTGCCGAGGCAGGGAAGCTAGTGGTAGTGCCCATGGATGGGAGCCA 144

Qy     129 ctggttcaccatgaggtcggtggtggagaaactcattctcaggggggcatgaggtggttgt 188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145 CTGGTTCACCATGAGGTCGGTGGTGGAGAAACTCATTCTCAGGGGGCATGAGGTGGTTGT 204

Qy     189 agtca-tgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaagactt 247
           ||||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     205 AGTCAGTGCCAGAGGTGAGTTGGCAACTGGGAAGATCAATGAATTGCACAGTGAAGACTT 264

Qy     248 attcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcccatgctc 307
```

```
DEFINITION  Homo sapiens, Similar to UDP glycosyltransferase 1 family,
            polypeptide A9, clone MGC:9353 IMAGE:3850654, mRNA, complete cds.
ACCESSION   BC020971
VERSION     BC020971.1  GI:18088086
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2423)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 20 Row: m Column: 8
            This clone was selected for full length sequencing because it
            passed the following selection criteria: GenomeScan gene
            prediction, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1. .2423
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:9353 IMAGE:3850654"
                     /tissue_type="Colon, adenocarcinoma"
                     /clone_lib="NIH_MGC_65"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             70. .1662
                     /codon_start=1
                     /product="Similar to UDP glycosyltransferase 1 family,
                     polypeptide A9"
                     /protein_id="AAH20971.1"
                     /db_xref="GI:18088087"
                     /translation="MARAGWTSPVPLCVCLLLTCGFAEAGKLLVVPMDGSHWFTMQSV
                     VEKLILRGHEVVVVMPEVSWQLERSLNCTVKTYSTSYTLEDQNREFMVFAHAQWKAQA
                     QSIFSLLMSSSSGFLDLFFSHCRSLFNDRKLVEYLKESSFDAVFLDPFDTCGLIVAKY
                     FSLPSVVFTRGIFCHHLEEGAQCPAPLSYVPNDLLGFSDAMTFKERVWNHIVHLEDHL
                     FCQYLFRNALEIASEILQTPVTAYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT      643 a    514 c    559 g    707 t
ORIGIN
```

```
  Query Match             94.6%;  Score 2194.6;  DB 9;  Length 2423;
  Best Local Similarity   97.0%;  Pred. No. 0;
  Matches 2247;  Conservative    0;  Mismatches   69;  Indels    1;  Gaps    1;

Qy       5 gctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtgtgtc 64
           ||||||||||||| |||||| || ||||||||||||||||| ||||| ||||||||||||
Db      56 GCTGCAGTTCTCTCATGGCTCGCGCAGGGTGGACCAGCCCCGTTCCTTTATGTGTGTGTC 115

Qy      65 tgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatggga 124
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     116 TACTGCTGACCTGTGGCTTTGCCGAGGCAGGGAAGCTGCTGGTAGTGCCCATGGATGGGA 175

Qy     125 gccactggttcaccatgaggtcggtggtggagaaactcattctcaggggggcatgaggtgg 184
           ||||||||||||||| |||||||||||||||||| || ||||||||||||||||||||||
Db     176 GTCACTGGTTCACCATGCAGTCGGTGGTGGAGAAACTTATCCTCAGGGGGCATGAGGTGG 235

Qy     185 ttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaaaga 244
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     236 TTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGAAAGATCACTGAATTGCACAGTGAAGA 295

Qy     245 cttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcccatg 304
           |||| ||||| || || || ||||| |||| | ||||||| |||| || ||| |||||||
Db     296 CTTACTCAACCTCGTACACTCTGGAAGATCAGAACCGGGAATTCATGGTTTTCGCCCATG 355

Qy     305 ctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgaca 364
           |||||||||||||||| | || |||||||| |||||||||||||| |||||| || ||
Db     356 CTCAATGGAAAGCACAGGCACAAAGTATATTTTCTCTATTAATGAGTTCATCCAGTGGTT 415

Qy     365 tttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaat 424
           || ||||||||||||||  |||||||||||||||||||| |||  ||||||||||||||
Db     416 TTCTTGACTTATTTTTTTCGCATTGCAGGAGTTTGTTTAATGACCGAAAATTAGTAGAAT 475

Qy     425 acttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggcttaa 484
           |||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||
Db     476 ACTTAAAGGAGAGTTCTTTTGATGCAGTGTTTCTGGATCCTTTTGATACCTGTGGCTTAA 535

Qy     485 ttgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgccact 544
           ||||||| |||||||||||||||||||| |||||||| |||||||||| ||||||||
Db     536 TTGTTGCTAAATATTTCTCCCTCCCCTCTGTGGTCTTCACCAGGGGAATATTTTGCCACC 595

Qy     545 atcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttag 604
           |||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||||
Db     596 ATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAATGATCTCTTAG 655

Qy     605 ggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggagg 664
           |||||||||||||||||||||||||||||||||||| |||||||||| ||||||||||||
Db     656 GGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTATGGAACCACATCGTGCACTTGGAGG 715

Qy     665 aacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaattctcc 724
           | |||||||| |||||   | |||| | ||||||||||||||||||||||||||||||||
Db     716 ACCATTTATTTTGCCAGTATCTTTTTAGAAATGCCCTAGAAATAGCCTCTGAAATTCTCC 775

Qy     725 aaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacgg 784
           |||| ||||| |||| |||||||||||||| ||||||||||||||||||||||||||||
Db     776 AAACCCCTGTCACGGCATATGATCTCTACAGTCACACATCAATTTGGTTGTTGCGAACGG 835

Qy     785 actttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatca 844
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     836 ACTTTGTTTTGGACTATCCCAAACCCGTGATGCCCAACATGATCTTCATTGGTGGTATCA 895

Qy     845 actgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctggag 904
           |||| ||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     896 ACTGTCATCAGGGAAAGCCATTGCCTATGGAATTTGAAGCCTACATTAATGCTTCTGGAG 955

Qy     905 aacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaagaaaag 964
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     956 AACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAG 1015

Qy     965 ctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactg 1024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1016 CTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTG 1075

Qy    1025 gaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacg 1084
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1076 GAACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACG 1135

Qy    1085 atctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgttt 1144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1136 ATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTT 1195

Qy    1145 atgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatgg 1204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    230 TTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGGAAAATCACTGAATTGCACAGTGAAGA 289

Qy    245 cttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcccatg 304
          |||| ||||| ||||| || |||||||||||||||||||| |||| || ||| ||| |||
Db    290 CTTACTCAACCTCATACACTCTGGAGGATCTGGACCGGGAATTCATGGATTTCGCCGATG 349

Qy    305 ctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgaca 364
          |||||||||||||||||||||||||| | | ||||||||| || |||||| |||||
Db    350 CTCAATGGAAAGCACAAGTACGAAGTTTGTTTTCTCTATTTCTGAGTTCATCCAATGGTT 409

Qy    365 tttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaat 424
          ||||| |||||||||||||  |||||||||||||||||||| ||| ||||||||||||||
Db    410 TTTTTAACTTATTTTTTTCGCATTGCAGGAGTTTGTTTAATGACCGAAAATTAGTAGAAT 469

Qy    425 acttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggcttaa 484
          |||||||||||||||||||||||||| |||||||| |||||||||||  ||||||||||
Db    470 ACTTAAAGGAGAGTTCTTTTGATGCGGTGTTTCTTGATCCTTTTGATGCCTGTGGCTTAA 529

Qy    485 ttgttgccaaatatttctccctccccctccgtggtcttcgccaggggaatactttgccact 544
          |||||||||||||||||||||||||||||| ||||||||||||||||||||  ||||||||
Db    530 TTGTTGCCAAATATTTCTCCCTCCCCTCTGTGGTCTTCGCCAGGGGAATAGCTTGCCACT 589

Qy    545 atcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttag 604
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    590 ATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTCTTAG 649

Qy    605 ggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggagg 664
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    650 GGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTACGGAACCACATCATGCACTTGGAGG 709

Qy    665 aacatttattatgccaccgtttttttcaaaaatgccctagaaatagcctctgaaattctcc 724
          |||||||||| |||||   ||||| ||||||||||||||||||||||||||||||||||||
Db    710 AACATTTATTTTGCCAGTATTTTTCCAAAAATGCCCTAGAAATAGCCTCTGAAATTCTCC 769

Qy    725 aaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacgg 784
          |||||||||| || |  |||||||||||||||||||||||||||||||||||||||| |
Db    770 AAACACCTGTCACAGCATATGATCTCTACAGCCACACATCAATTTGGTTGTTGCGAACAG 829

Qy    785 actttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatca 844
          ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    830 ACTTTGTTTTGGACTATCCCAAACCCGTGATGCCCAATATGATCTTCATTGGTGGTATCA 889

Qy    845 actgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctggag 904
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    890 ACTGCCATCAGGGAAAGCCATTGCCTATGGAATTTGAAGCCTACATTAATGCTTCTGGAG 949

Qy    905 aacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaagaaag 964
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    950 AACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAG 1009

Qy    965 ctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactg 1024
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1010 CTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTG 1069

Qy   1025 gaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacg 1084
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1070 GAACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACG 1129

Qy   1085 atctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgttt 1144
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1130 ATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTT 1189

Qy   1145 atgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatgg 1204
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1190 ATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGG 1249

Qy   1205 acaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatga 1264
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1250 ACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGA 1309

Qy   1265 cttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggaga 1324
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1310 CTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGA 1369
```

```
Qy   1325 acatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccg 1384
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1370 ACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCG 1429

Qy   1385 tgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagccc 1444
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1430 TGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCC 1489

Qy   1445 acgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcg 1504
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1490 ACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCG 1549

Qy   1505 tgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttgg 1564
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1550 TGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGG 1609

Qy   1565 ggaaaaaagtggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtggga 1624
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1610 GGAAAAAAGTGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGA 1669

Qy   1625 aataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtg 1684
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1670 AATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTG 1729

Qy   1685 ttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtg 1744
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1730 TTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTG 1789

Qy   1745 ctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaat 1804
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1790 CTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAAT 1849

Qy   1805 atgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggac 1864
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1850 ATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGAC 1909

Qy   1865 gtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttggga 1924
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1910 GTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGA 1969

Qy   1925 gtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatctta 1984
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1970 GTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTA 2029

Qy   1985 tcttttggcttctgcagatggttgcaattgatc-cttaaccaataatggtcagtcctcat 2043
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   2030 TCTTTTGGCTTCTGCAGATGGTTGCAATTGATCTCTTAACCAATAATGGTCAGTCCTCAT 2089

Qy   2044 ctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacctttt 2103
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2090 CTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTT 2149

Qy   2104 agagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttctt 2163
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   2150 AGAGTGTACGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTTTTT 2209

Qy   2164 gctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatc 2223
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2210 GCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATAATC 2269

Qy   2224 attgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgct 2283
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2270 ATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGTGCT 2329

Qy   2284 gggcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||||||||||| |||||||||||| |||||
Db   2330 GGGCAAGTTTACTTTTTTTTTGATGTTTCCTCCAACT 2366
```

RESULT 6
BC020971
LOCUS BC020971 2423 bp mRNA linear PRI 22-JAN-2002

```
Qy   1205 acaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatga 1264
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1220 ACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGA 1279

Qy   1265 cttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggaga 1324
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1280 CTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGA 1339

Qy   1325 acatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccg 1384
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1340 ACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCCGGTGGAGCCGCTGGACCTGGCCG 1399

Qy   1385 tgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagccc 1444
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1400 TGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCC 1459

Qy   1445 acgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcg 1504
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1460 ACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCG 1519

Qy   1505 tgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttgg 1564
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1520 TGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGG 1579

Qy   1565 ggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtggga 1624
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1580 GGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGA 1639

Qy   1625 aataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtg 1684
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1640 AATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTG 1699

Qy   1685 ttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtg 1744
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1700 TTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTG 1759

Qy   1745 ctttaaaaaattctctaaataaaaataatagactcgctagtcagtaaagatatttgaat 1804
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1760 CTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAAT 1819

Qy   1805 atgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggac 1864
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1820 ATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGAC 1879

Qy   1865 gtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttggga 1924
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1880 GTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGA 1939

Qy   1925 gtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatctta 1984
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1940 GTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTA 1999

Qy   1985 tcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatc 2044
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2000 TCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCTCATC 2059

Qy   2045 tctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtaccttta 2104
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2060 TCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTTA 2119

Qy   2105 gagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttcttg 2164
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   2120 GAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTTTTTG 2179

Qy   2165 ctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatca 2224
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2180 CTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATAATCA 2239

Qy   2225 ttgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgctg 2284
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2240 TTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGTGCTG 2299

Qy   2285 ggcaagtttactttttttctgatgtttcctacaact 2320
          ||||||||||||||||||| |||||||||| |||||
Db   2300 GGCAAGTTTACTTTTTTTTTGATGTTTCCTCCAACT 2335
```

RESULT 5
AF462268
LOCUS AF462268 2394 bp mRNA linear PRI 07-FEB-2002
DEFINITION Homo sapiens UDP-glucuronosyltransferase UGT1A8*1 (UGT1A8) mRNA, UGT1A8*1 allele, complete cds.
ACCESSION AF462268
VERSION AF462268.1 GI:18568218
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2394)
AUTHORS Huang,Y.-H., Nguyen,N., Geske,D. and Tukey,R.
TITLE Identification and functional characterization of UGT1A8*1, UGT1A8*2 and UGT1A8*3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2394)
AUTHORS Huang,Y.-H., Nguyen,N., Geske,D. and Tukey,R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2001) Pharmacology, UCSD, 9500 Gilman Drive, La Jolla, CA 92093, USA
FEATURES Location/Qualifiers
source 1. .2394
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="colon"
gene 1. .2394
/gene="UGT1A8"
/allele="UGT1A8*1"
CDS 64. .1656
/gene="UGT1A8"
/codon_start=1
/product="UDP-glucuronosyltransferase UGT1A8*1"
/protein_id="AAL75964.1"
/db_xref="GI:18568219"
/translation="MARTGWTSPIPLCVSLLLTCGFAEAGKLLVVPMDGSHWFTMQSV
VEKLILRGHEVVVVMPEVSWQLGKSLNCTVKTYSTSYTLEDLDREFMDFADAQWKAQV
RSLFSLFLSSSNGFFNLFFSHCRSLFNDRKLVEYLKESSFDAVFLDPFDACGLIVAKY
FSLPSVVFARGIACHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHL
FCQYFSKNALEIASEILQTPVTAYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
YGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT 624 a 513 c 555 g 702 t
ORIGIN

Query Match 95.1%; Score 2207.4; DB 9; Length 2394;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 2255; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

```
Qy      5 gctgcagttctctgatggcttgcacagggtggaccagcccccttcctctatgtgtgtgtc 64
          |||||||||||||| |||||| |||||||||||||||||||| |||| |||||||| | ||
Db     50 GCTGCAGTTCTCTCATGGCTCGCACAGGGTGGACCAGCCCCATTCCCCTATGTGTTTCTC 109

Qy     65 tgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatggga 124
          |||||||||||||||||||||| |||||||||||||| ||||||||||||||||||||||
Db    110 TGCTGCTGACCTGTGGCTTTGCTGAGGCAGGGAAGCTGCTGGTAGTGCCCATGGATGGGA 169

Qy    125 gccactggttcaccatgaggtcggtggtggagaaactcattctcagggggcatgaggtgg 184
          | ||||||||||||||  |||||||||||||||||| || ||||||||||||||||||||
Db    170 GTCACTGGTTCACCATGCAGTCGGTGGTGGAGAAACTTATCCTCAGGGGGCATGAGGTGG 229

Qy    185 ttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaaga 244
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
```

```
Qy   2161 cttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagata 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATA 2220

Qy   2221 atcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagt 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 ATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGT 2280

Qy   2281 gctgggcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||||||
Db   2281 GCTGGGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 2320


RESULT   4
AF462267
LOCUS       AF462267                2363 bp    mRNA    linear   PRI 07-FEB-2002
DEFINITION  Homo sapiens UDP-glucuronosyltransferase UGT1A8*2 (UGT1A8) mRNA,
            UGT1A8*2 allele, complete cds.
ACCESSION   AF462267
VERSION     AF462267.1  GI:18568216
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2363)
  AUTHORS   Huang,Y.-H., Geske,D., Nguyen,N. and Tukey,R.
  TITLE     Identification and functional characterization of UGT1A8*1,
            UGT1A8*2 and UGT1A8*3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2363)
  AUTHORS   Huang,Y.-H., Geske,D., Nguyen,N. and Tukey,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2001) Pharmacology, UCSD, 9500 Gilman Drive, La
            Jolla, CA 92093, USA
FEATURES             Location/Qualifiers
     source          1. .2363
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="colon"
     gene            1. .2363
                     /gene="UGT1A8"
                     /allele="UGT1A8*2"
     CDS             34. .1626
                     /gene="UGT1A8"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase UGT1A8*2"
                     /protein_id="AAL75963.1"
                     /db_xref="GI:18568217"
                     /translation="MARTGWTSPIPLCVSLLLTCGFAEAGKLLVVPMDGSHWFTMQSV
                     VEKLILRGHEVVVVMPEVSWQLGKSLNCTVKTYSTSYTLEDLDREFMDFADAQWKAQV
                     RSLFSLFLSSSNGFFNLFFSHCRSLFNDRKLVEYLKESSFDAVFLDPFDACGLIVAKY
                     FSLPSVVFARGIGCHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHL
                     FCQYFSKNALEIASEILQTPVTAYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT      617 a    503 c    551 g    692 t
ORIGIN


  Query Match           95.6%;  Score 2218.4;  DB 9;  Length 2363;
  Best Local Similarity  97.4%;  Pred. No. 0;
  Matches 2255;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy       5 gctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtgtgtc 64
          |||||||||||| |||||| |||||||||||||||||| |||| |||||||| | ||
Db      20 GCTGCAGTTCTCTCATGGCTCGCACAGGGTGGACCAGCCCCATTCCCCTATGTGTTTCTC 79

Qy      65 tgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatggga 124
           |||| |||||||||||||||| |||||||||||||| ||||||||||||||||||||||

Db      80 TGCTGCTGACCTGTGGCTTTGCTGAGGCAGGGAAGCTGCTGGTAGTGCCCATGGATGGGA 139

Qy     125 gccactggttcaccatgaggtcggtggtggagaaactcattctcaggggggcatgaggtgg 184
           | |||||||||||||||  |||||||||||||||||| || ||||||||||||||||||
Db     140 GTCACTGGTTCACCATGCAGTCGGTGGTGGAGAAACTTATCCTCAGGGGGCATGAGGTGG 199

Qy     185 ttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaaga 244
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     200 TTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGGAAAATCACTGAATTGCACAGTGAAGA 259

Qy     245 cttattcaacttcatatacccctggaggatctggaccgggagttcaaggcttttgcccatg 304
           |||| ||||| ||||| || ||||||||||||||||||||| |||| || ||| ||| |||
Db     260 CTTACTCAACCTCATACACTCTGGAGGATCTGGACCGGGAATTCATGGATTTCGCCGATG 319

Qy     305 ctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgaca 364
           |||||||||||||||||||||||||| | | ||||||||  || ||||||  |||||
Db     320 CTCAATGGAAAGCACAAGTACGAAGTTTGTTTTCTCTATTTCTGAGTTCATCCAATGGTT 379

Qy     365 tttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaat 424
           |||||  |||||||||||||  |||||||||||||||||||| |||  ||||||||||||||
Db     380 TTTTTAACTTATTTTTTTCGCATTGCAGGAGTTTGTTTAATGACCGAAAATTAGTAGAAT 439

Qy     425 acttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggcttaa 484
           |||||||||||||||||||||||||||||| |||||||||  ||||||||||||||||||
Db     440 ACTTAAAGGAGAGTTCTTTTGATGCGGTGTTTCTTGATCCTTTTGATGCCTGTGGCTTAA 499

Qy     485 ttgttgccaaatatttctccctccccctccgtggtcttcgccaggggaatactttgccact 544
           |||||||||||||||||||||||||||||| ||||||||||||||||||  ||||||||
Db     500 TTGTTGCCAAATATTTCTCCCTCCCCTCTGTGGTCTTCGCCAGGGGAATAGGTTGCCACT 559

Qy     545 atcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttag 604
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     560 ATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTCTTAG 619

Qy     605 ggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggagg 664
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     620 GGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTACGGAACCACATCATGCACTTGGAGG 679

Qy     665 aacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaattctcc 724
           |||||||||| |||||   ||||| ||||||||||||||||||||||||||||||||||||
Db     680 AACATTTATTTTGCCAGTATTTTTCCAAAAATGCCCTAGAAATAGCCTCTGAAATTCTCC 739

Qy     725 aaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacgg 784
           |||||||||| || | ||||||||||||||||||||||||||||||||||||||||||| |
Db     740 AAACACCTGTCACAGCATATGATCTCTACAGCCACACATCAATTTGGTTGTTGCGAACAG 799

Qy     785 actttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatca 844
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     800 ACTTTGTTTTGGACTATCCCAAACCCGTGATGCCCAATATGATCTTCATTGGTGGTATCA 859

Qy     845 actgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctggag 904
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     860 ACTGCCATCAGGGAAAGCCATTGCCTATGGAATTTGAAGCCTACATTAATGCTTCTGGAG 919

Qy     905 aacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaagaaag 964
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     920 AACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAG 979

Qy     965 ctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactg 1024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     980 CTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTG 1039

Qy    1025 gaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacg 1084
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1040 GAACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACG 1099

Qy    1085 atctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgttt 1144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1100 ATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTT 1159

Qy    1145 atgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatgg 1204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1160 ATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGG 1219
```

```
  Matches 2320;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTCAGCTGCAGTTCTCTGATGGCTTGCACAGGGTGGACCAGCCCCCTTCCTCTATGTGTG 60

Qy      61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGTCTGCTGCTGACCTGTGGCTTTGCCGAGGCAGGGAAGCTACTGGTAGTGCCCATGGAT 120

Qy     121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggggcatgag 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGGAGCCACTGGTTCACCATGAGGTCGGTGGTGGAGAAACTCATTCTCAGGGGGCATGAG 180

Qy     181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GTGGTTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGGAAGATCACTGAATTGCACAGTG 240

Qy     241 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AAGACTTATTCAACTTCATATACCCTGGAGGATCTGGACCGGGAGTTCAAGGCTTTTGCC 300

Qy     301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CATGCTCAATGGAAAGCACAAGTACGAAGTATATATTCTCTATTAATGGGTTCATACAAT 360

Qy     361 gacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagta 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GACATTTTTGACTTATTTTTTTCAAATTGCAGGAGTTTGTTTAAAGACAAAAAATTAGTA 420

Qy     421 gaatacttaaaggagagttctttttgatgcagtgtttctcgatccttttgataactgtggc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GAATACTTAAAGGAGAGTTCTTTTGATGCAGTGTTTCTCGATCCTTTTGATAACTGTGGC 480

Qy     481 ttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TTAATTGTTGCCAAATATTTCTCCCTCCCCTCCGTGGTCTTCGCCAGGGGAATACTTTGC 540

Qy     541 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CACTATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTC 600

Qy     601 ttagggttctcagatgccatgacttttcaaggagagagtacggaaccacatcatgcacttg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TTAGGGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTACGGAACCACATCATGCACTTG 660

Qy     661 gaggaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaatt 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GAGGAACATTTATTATGCCACCGTTTTTTCAAAAATGCCCTAGAAATAGCCTCTGAAATT 720

Qy     721 ctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcga 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CTCCAAACACCTGTTACGGAGTATGATCTCTACAGCCACACATCAATTTGGTTGTTGCGA 780

Qy     781 acggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggt 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 ACGGACTTTGTTTTGGACTATCCCAAACCCGTGATGCCCAACATGATCTTCATTGGTGGT 840

Qy     841 atcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttct 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ATCAACTGCCATCAGGGAAAGCCGTTGCCTATGGAATTTGAAGCCTACATTAATGCTTCT 900

Qy     901 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GGAGAACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAG 960

Qy     961 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 AAAGCTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTAC 1020

Qy    1021 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaa 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 ACTGGAACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAA 1080

Qy    1081 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 AACGATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGT 1140

Qy    1141 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAG 1200

Qy    1201 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ATGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAA 1260

Qy    1261 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 ATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAG 1320

Qy    1321 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 GAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTG 1380

Qy    1381 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 GCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCA 1440

Qy    1441 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 GCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCC 1500

Qy    1501 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 GTCGTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGC 1560

Qy    1561 ttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggt 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 TTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGT 1620

Qy    1621 gggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatc 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 GGGAAATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATC 1680

Qy    1681 agtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccag 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 AGTGTTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAG 1740

Qy    1741 agtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatattt 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 AGTGCTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTT 1800

Qy    1801 gaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcaga 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 GAATATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGA 1860

Qy    1861 ggacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgttt 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 GGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTT 1920

Qy    1921 gggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaat 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 GGGAGTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAAT 1980

Qy    1981 cttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcct 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 CTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCT 2040

Qy    2041 catctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacc 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 CATCTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACC 2100

Qy    2101 tttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgattt 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 TTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTT 2160
```

```
Db    1021 ACTCGAACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAA 1080

Qy    1081 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 AACGATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGT 1140

Qy    1141 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAG 1200

Qy    1201 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ATGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAA 1260

Qy    1261 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 ATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAG 1320

Qy    1321 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 GAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTG 1380

Qy    1381 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 GCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCA 1440

Qy    1441 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 GCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCC 1500

Qy    1501 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 GTCGTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGC 1560

Qy    1561 ttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggt 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 TTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGT 1620

Qy    1621 gggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatc 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 GGGAAATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATC 1680

Qy    1681 agtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccag 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 AGTGTTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAG 1740

Qy    1741 agtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatattt 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 AGTGCTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTT 1800

Qy    1801 gaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcaga 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 GAATATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGA 1860

Qy    1861 ggacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgttt 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 GGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTT 1920

Qy    1921 gggagtgcgggattcaaaggtggtcccaccgctgccctactgcaaatggcagttttaat 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 GGGAGTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAAT 1980

Qy    1981 cttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcct 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 CTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCT 2040

Qy    2041 catcctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacc 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 CATCTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACC 2100

Qy    2101 tttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgattt 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 TTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTT 2160

Qy    2161 cttgctttgggggaaaaagaaatgatgctatgaaattggtgggtggtgtatttgagaagata 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 CTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATA 2220

Qy    2221 atcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagt 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 ATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGT 2280

Qy    2281 gctgggcaagtttactttttttctgatgtttcctacaact 2320
           ||||||||||||||||||||||||||||||||||||||||
Db    2281 GCTGGGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 2320
```

RESULT 3
AF056188
LOCUS AF056188 2320 bp mRNA linear PRI 13-AUG-1998
DEFINITION Homo sapiens UDP-glucuronosyltransferase 1A9 (UGT1A9) mRNA, complete cds.
ACCESSION AF056188
VERSION AF056188.1 GI:3414797
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2320)
AUTHORS Ciotti,M., Potter,C. and Owens,I.S.
TITLE Human phenol metabolizing UDP-glucuronosyltransferase
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2320)
AUTHORS Ciotti,M., Potter,C. and Owens,I.S.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-1998) Heritable Disorders Branch, NICHD, 9000 Rockville Pike, Bethesda, MD 20892, USA
REFERENCE 3 (bases 1 to 2320)
AUTHORS Ciotti,M., Potter,C. and Owens,I.S.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1998) Heritable Disorders Branch, NICHD, 9000 Rockville Pike, Bethesda, MD 20892, USA
REMARK Sequence update by submitter
COMMENT On Aug 13, 1998 this sequence version replaced gi:3025895.
FEATURES Location/Qualifiers
source 1. .2320
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q37"
/tissue_type="liver"
gene 1. .2320
/gene="UGT1A9"
CDS 19. .1611
/gene="UGT1A9"
/codon_start=1
/product="UDP-glucuronosyltransferase 1A9"
/protein_id="AAC31425.1"
/db_xref="GI:3025896"
/translation="MACTGWTSPLPLCVCLLLTCGFAEAGKLLVVPMDGSHWFTMRSV
VEKLILRGHEVVVVMPEVSWQLGRSLNCTVKTYSTSYTLEDLDREFKAFAHAQWKAQV
RSIYSLLMGSYNDIFDLFFSNCRSLFKDKKLVEYLKESSFDAVFLDPFDNCGLIVAKY
FSLPSVVFARGILCHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHL
LCHRFFKNALEIASEILQTPVTEYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
YGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT 608 a 495 c 547 g 670 t
ORIGIN

Query Match 100.0%; Score 2320; DB 9; Length 2320;
Best Local Similarity 100.0%; Pred. No. 0;

```
Db   1861 GGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTT 1920

Qy   1921 gggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaat 1980

Db   1921 GGGAGTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAAT 1980

Qy   1981 cttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcct 2040

Db   1981 CTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCT 2040

Qy   2041 catctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacc 2100

Db   2041 CATCTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACC 2100

Qy   2101 tttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgattt 2160

Db   2101 TTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTT 2160

Qy   2161 cttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagata 2220

Db   2161 CTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATA 2220

Qy   2221 atcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagt 2280

Db   2221 ATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGT 2280

Qy   2281 gctgggcaagtttacttttttttctgatgtttcctacaact 2320

Db   2281 GCTGGGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 2320
```

```
RESULT   2
AX286764
LOCUS       AX286764               2320 bp    DNA     linear   PAT 21-NOV-2001
DEFINITION  Sequence 3 from Patent WO0180896.
ACCESSION   AX286764
VERSION     AX286764.1  GI:17048799
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ratain,M.J., Innocenti,F. and Iyer,L.
  TITLE     Flavopiridol drug combinations and methods with reduced side
            effects
  JOURNAL   Patent: WO 0180896-A 3 01-NOV-2001;
            ARCH DEVELOPMENT CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2320
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             19. .1611
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD12275.1"
                     /db_xref="GI:17048800"
                     /translation="MACTGWTSPLPLCVCLLLTCGFAEAGKLLVVPMDGSHWFTMRSV
                     VEKLILRGHEVVVVMPEVSWQLGRSLNCTVKTYSTSYTLEDLDREFKAFAHAQWKAQV
                     RSIYSLLMGSYNDIFDLFFSNCRSLFKDKKLVEYLKESSFDAVFLDPFDNCGLIVAKY
                     FSLPSVVFARGILCHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHL
                     LCHRFFKNALEIASEILQTPVTEYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT      608 a    495 c    547 g    670 t
ORIGIN
```

Query Match 100.0%; Score 2320; DB 6; Length 2320;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 2320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60

Db      1 CTCAGCTGCAGTTCTCTGATGGCTTGCACAGGGTGGACCAGCCCCCTTCCTCTATGTGTG 60

Qy     61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120

Db     61 TGTCTGCTGCTGACCTGTGGCTTTGCCGAGGCAGGGAAGCTACTGGTAGTGCCCATGGAT 120

Qy    121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgag 180

Db    121 GGGAGCCACTGGTTCACCATGAGGTCGGTGGTGGAGAAACTCATTCTCAGGGGGCATGAG 180

Qy    181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240

Db    181 GTGGTTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGGAAGATCACTGAATTGCACAGTG 240

Qy    241 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 300

Db    241 AAGACTTATTCAACTTCATATACCCTGGAGGATCTGGACCGGGAGTTCAAGGCTTTTGCC 300

Qy    301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360

Db    301 CATGCTCAATGGAAAGCACAAGTACGAAGTATATATTCTCTATTAATGGGTTCATACAAT 360

Qy    361 gacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagta 420

Db    361 GACATTTTTGACTTATTTTTTTCAAATTGCAGGAGTTTGTTTAAAGACAAAAAATTAGTA 420

Qy    421 gaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggc 480

Db    421 GAATACTTAAAGGAGAGTTCTTTTGATGCAGTGTTTCTCGATCCTTTTGATAACTGTGGC 480

Qy    481 ttaattgttgccaaatatttctccctccccctccgtggtcttcgccaggggaatactttgc 540

Db    481 TTAATTGTTGCCAAATATTTCTCCCTCCCCTCCGTGGTCTTCGCCAGGGGAATACTTTGC 540

Qy    541 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 600

Db    541 CACTATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTC 600

Qy    601 ttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttg 660

Db    601 TTAGGGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTACGGAACCACATCATGCACTTG 660

Qy    661 gaggaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaatt 720

Db    661 GAGGAACATTTATTATGCCACCGTTTTTTCAAAAATGCCCTAGAAATAGCCTCTGAAATT 720

Qy    721 ctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcga 780

Db    721 CTCCAAACACCTGTTACGGAGTATGATCTCTACAGCCACACATCAATTTGGTTGTTGCGA 780

Qy    781 acggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggt 840

Db    781 ACGGACTTTGTTTTGGACTATCCCAAACCCGTGATGCCCAACATGATCTTCATTGGTGGT 840

Qy    841 atcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttct 900

Db    841 ATCAACTGCCATCAGGGAAAGCCGTTGCCTATGGAATTTGAAGCCTACATTAATGCTTCT 900

Qy    901 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 960

Db    901 GGAGAACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAG 960

Qy    961 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1020

Db    961 AAAGCTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTAC 1020

Qy   1021 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaa 1080
```

```
                     /codon_start=1
                     /protein_id="CAD12274.1"
                     /db_xref="GI:17048798"
                     /translation="MACTGWTSPLPLCVCLLLTCGFAEAGKLLVVPMDGSHWFTMRSV
                     VEKLILRGHEVVVVMPEVSWQLGRSLNCTVKTYSTSYTLEDLDREFKAFAHAQWKAQV
                     RSIYSLLMGSYNDIFDLFFSNCRSLFKDKKLVEYLKESSFDAVFLDPFDNCGLIVAKY
                     FSLPSVVFARGILCHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHL
                     LCHRFFKNALEIASEILQTPVTEYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT      608 a    495 c    547 g    670 t
ORIGIN
```

Query Match 100.0%; Score 2320; DB 6; Length 2320;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60
Db       1 CTCAGCTGCAGTTCTCTGATGGCTTGCACAGGGTGGACCAGCCCCCTTCCTCTATGTGTG 60

Qy      61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120
Db      61 TGTCTGCTGCTGACCTGTGGCTTTGCCGAGGCAGGGAAGCTACTGGTAGTGCCCATGGAT 120

Qy     121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggggcatgag 180
Db     121 GGGAGCCACTGGTTCACCATGAGGTCGGTGGTGGAGAAACTCATTCTCAGGGGGCATGAG 180

Qy     181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240
Db     181 GTGGTTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGGAAGATCACTGAATTGCACAGTG 240

Qy     241 aagacttattcaacttcatatacccctggaggatctggaccgggagttcaaggcttttgcc 300
Db     241 AAGACTTATTCAACTTCATATACCCTGGAGGATCTGGACCGGGAGTTCAAGGCTTTTGCC 300

Qy     301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360
Db     301 CATGCTCAATGGAAAGCACAAGTACGAAGTATATATTCTCTATTAATGGGTTCATACAAT 360

Qy     361 gacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagta 420
Db     361 GACATTTTTGACTTATTTTTTTCAAATTGCAGGAGTTTGTTTAAAGACAAAAAATTAGTA 420

Qy     421 gaatacttaaaggagagttctttgatgcagtgtttctcgatcctttgataactgtggc 480
Db     421 GAATACTTAAAGGAGAGTTCTTTTGATGCAGTGTTTCTCGATCCTTTTGATAACTGTGGC 480

Qy     481 ttaattgttgccaaatatttctccctccccctccgtggtcttcgccaggggaatactttgc 540
Db     481 TTAATTGTTGCCAAATATTTCTCCCTCCCCTCCGTGGTCTTCGCCAGGGGAATACTTTGC 540

Qy     541 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 600
Db     541 CACTATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTC 600

Qy     601 ttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttg 660
Db     601 TTAGGGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTACGGAACCACATCATGCACTTG 660

Qy     661 gaggaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaatt 720
Db     661 GAGGAACATTTATTATGCCACCGTTTTTTCAAAAATGCCCTAGAAATAGCCTCTGAAATT 720

Qy     721 ctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcga 780
Db     721 CTCCAAACACCTGTTACGGAGTATGATCTCTACAGCCACACATCAATTTGGTTGTTGCGA 780

Qy     781 acggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggt 840
Db     781 ACGGACTTTGTTTTGGACTATCCCAAACCCGTGATGCCCAACATGATCTTCATTGGTGGT 840

Qy     841 atcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttct 900
Db     841 ATCAACTGCCATCAGGGAAAGCCGTTGCCTATGGAATTTGAAGCCTACATTAATGCTTCT 900

Qy     901 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 960
Db     901 GGAGAACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAG 960

Qy     961 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1020
Db     961 AAAGCTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTAC 1020

Qy    1021 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaa 1080
Db    1021 ACTGGAACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAA 1080

Qy    1081 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1140
Db    1081 AACGATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGT 1140

Qy    1141 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1200
Db    1141 GTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAG 1200

Qy    1201 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1260
Db    1201 ATGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAA 1260

Qy    1261 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1320
Db    1261 ATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAG 1320

Qy    1321 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1380
Db    1321 GAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTG 1380

Qy    1381 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1440
Db    1381 GCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCA 1440

Qy    1441 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1500
Db    1441 GCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCC 1500

Qy    1501 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1560
Db    1501 GTCGTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGC 1560

Qy    1561 ttggggaaaaaaggggcgagttaagaaagcccacaaatccaagacccattgagaagtgggt 1620
Db    1561 TTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGT 1620

Qy    1621 gggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatc 1680
Db    1621 GGGAAATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATC 1680

Qy    1681 agtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccag 1740
Db    1681 AGTGTTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAG 1740

Qy    1741 agtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatattt 1800
Db    1741 AGTGCTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTT 1800

Qy    1801 gaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcaga 1860
Db    1801 GAATATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGA 1860

Qy    1861 ggacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgttt 1920
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 21, 2002, 11:57:31 ; Search time 8248.67 Seconds
(without alignments)
5885.745 Million cell updates/sec

Title: US-09-835-082-3
Perfect score: 2320
Sequence: 1 ctcagctgcagttctctgat.....ttctgatgtttcctacaact 2320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2320	100.0	2320	6	AX286762	AX286762 Sequence
	2	2320	100.0	2320	6	AX286764	AX286764 Sequence
	3	2320	100.0	2320	9	AF056188	AF056188 Homo sapi
	4	2218.4	95.6	2363	9	AF462267	AF462267 Homo sapi
	5	2207.4	95.1	2394	9	AF462268	AF462268 Homo sapi
	6	2194.6	94.6	2423	9	BC020971	BC020971 Homo sapi
	7	2156.8	93.0	2392	9	S55985	S55985 Homo sapien
	8	2076	89.5	2396	9	AF104336	AF104336 Macaca fa
	9	1895.8	81.7	2003	9	AK025403	AK025403 Homo sapi
	10	1636.6	70.5	2368	9	HUMHUGBR2	M57951 Human bilir
	11	1629	70.2	2388	9	BC019861	BC019861 Homo sapi
	12	1610.8	69.4	2351	9	HUMHUGBR1	M57899 Human bilir
	13	1539.2	66.3	2422	6	AX336711	AX336711 Sequence
	14	1539.2	66.3	2422	9	HUMUGT1FA	J04093 Homo sapien
	15	1493.8	64.4	1593	9	AF030310	AF030310 Homo sapi
	16	1484.2	64.0	1593	9	HSU89507	U89507 Human UDP-g
	17	1477	63.7	1621	9	AF104338	AF104338 Macaca fa
	18	1470.6	63.4	2440	9	AF104337	AF104337 Macaca fa
	19	1463.6	63.1	1593	9	HSU89508	U89508 Human UDP-g
	20	1157.8	49.9	2241	10	AF461735	AF461735 Rattus no
	21	1129.4	48.7	2203	10	BC019434	BC019434 Mus muscu
	22	1123.4	48.4	1590	4	AF011902	AF011902 Oryctolag
	23	1110.4	47.9	2155	10	MUSUGTP	L27122 Mus musculu
	24	1105.2	47.6	2332	4	AB018478	AB018478 Ovis arie
	25	1105	47.6	2282	10	AF461738	AF461738 Rattus no
	26	1004.4	43.3	1561	9	HUMUGT03	M84124 Human UDP-g
c	27	1004.4	43.3	68770	9	AC006985	AC006985 Homo sapi
c	28	1004.4	43.3	176619	2	AC026497	AC026497 Homo sapi
	29	1004.4	43.3	198872	9	AF297093	AF297093 Homo sapi
	30	937	40.4	2301	10	RNU75903	U75903 Rattus norv
	31	878.6	37.9	1255	9	AF297091	AF297091 Homo sapi
c	32	878.6	37.9	148438	9	AC019072	AC019072 Homo sapi
	33	873	37.6	1602	9	AF360121	AF360121 Macaca mu
	34	872.2	37.6	1647	9	AF104339	AF104339 Macaca fa
	35	868.2	37.4	1602	9	AY029169	AY029169 Macaca mu
	36	819.8	35.3	2816	4	AF039138	AF039138 Felis cat
c	37	799.8	34.5	183826	2	AC091778	AC091778 Papio cyn
	38	794	34.2	2300	10	MMU16818	U16818 Mus musculu
	39	788.4	34.0	2037	4	OCU09101	U09101 Oryctolagus
	40	783.2	33.8	1765	10	RATUDPGTA	M34007 Rat bilirub
	41	781.8	33.7	2343	10	AF461737	AF461737 Rattus no
	42	779	33.6	1244	9	HSU42604	U42604 Human UDP-g
	43	777.8	33.5	1927	10	RATUDPGT	J02612 Rat UDP-glu
	44	774	33.4	2239	10	RNU20551	U20551 Rattus norv
	45	772.2	33.3	2484	10	RATUGT	J05132 Rat 3-methy

ALIGNMENTS

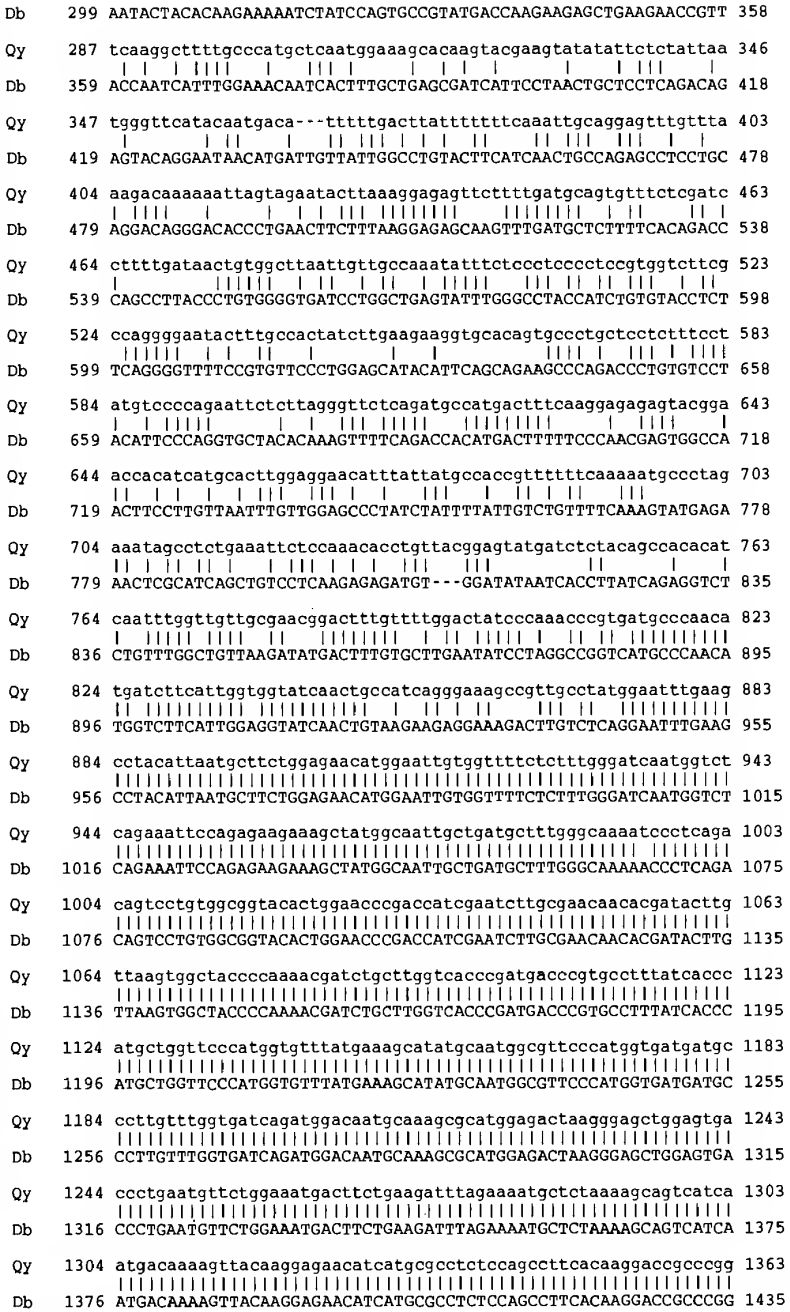
RESULT 1
AX286762
LOCUS AX286762 2320 bp DNA linear PAT 21-NOV-2001
DEFINITION Sequence 1 from Patent WO0180896.
ACCESSION AX286762
VERSION AX286762.1 GI:17048797
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Ratain,M.J., Innocenti,F. and Iyer,L.
TITLE Flavopiridol drug combinations and methods with reduced side effects
JOURNAL Patent: WO 0180896-A 1 01-NOV-2001;
ARCH DEVELOPMENT CORPORATION (US)
FEATURES Location/Qualifiers
source 1. .2320
/organism="Homo sapiens"
/db_xref="taxon:9606"
CDS 19. .1611
/note="unnamed protein product"

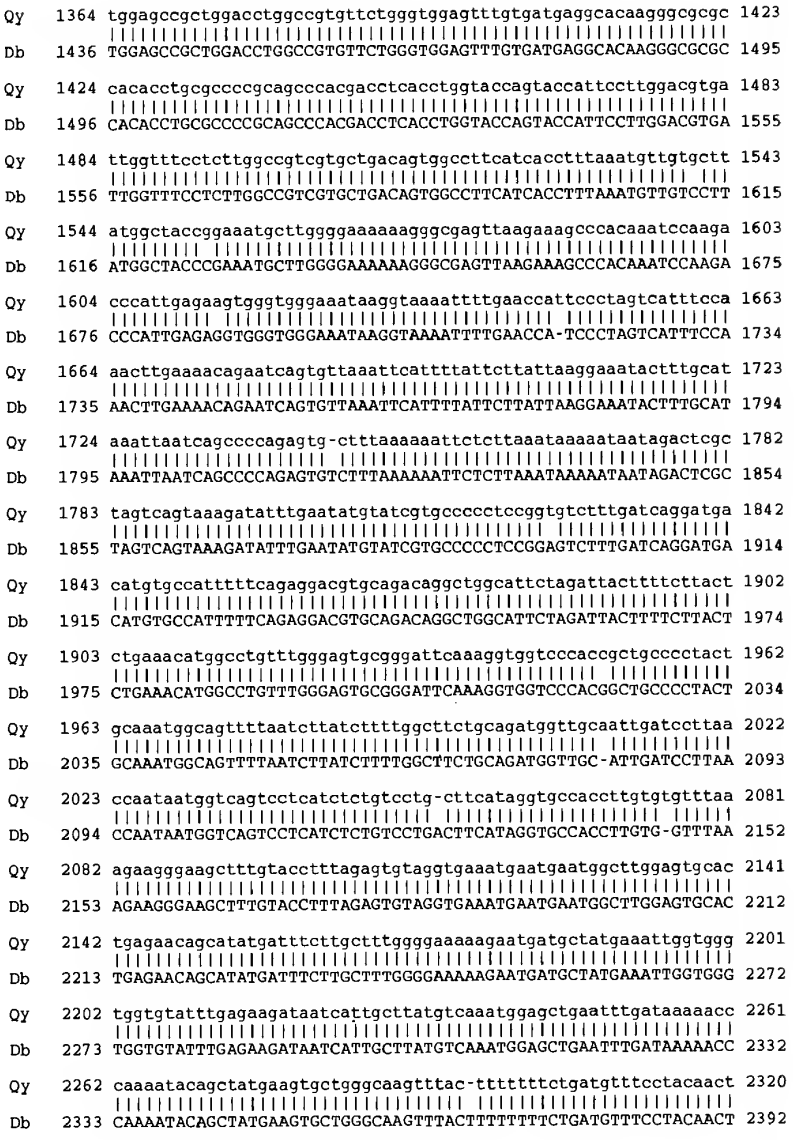
```
Qy   1459 taccagtaccattccttggacgtgattggtttcctcttggccgtcgtgctgacagtggcc 1518
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGCC 1500

Qy   1519 ttcatcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcga 1578
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGGAAAAAAGGGCGA 1560

Qy   1579 gttaagaaagcccacaaatccaagacccattga 1611
          ||||||||||||||||||||||||||||||||
Db   1561 GTTAAGAAAGCCCACAAATCCAAGACCCATTGA 1593
```

Search completed: August 21, 2002, 11:57:31
Job time: 32834 sec

```
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1593)
  AUTHORS   Strassburg,C.P., Manns,M.P. and Tukey,R.H.
  TITLE     Expression of the UDP-glucuronosyltransferase 1A locus in human
            colon. Identification and characterization of the novel
            extrahepatic UGT1A8
  JOURNAL   J. Biol. Chem. 273 (15), 8719-8726 (1998)
  MEDLINE   98204857
   PUBMED   9535849
REFERENCE   2  (bases 1 to 1593)
  AUTHORS   Strassburg,C.P. and Tukey,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1997) Pharmacology and UCSD Cancer Center,
            University of California, San Diego, 9500 Gilman Dr, BSB 4th Floor
            Rm 4021, La Jolla, CA 92093-0636, USA
FEATURES             Location/Qualifiers
     source          1. .1593
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
     gene            1. .1593
                     /gene="UGT1A8"
     CDS             1. .1593
                     /gene="UGT1A8"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A8"
                     /protein_id="AAB84259.1"
                     /db_xref="GI:2613044"
                     /translation="MARTGWTSPIPLCVSLLLTCGFAEAGKLLVVPMDGSHWFTMQSV
                     VEKLILRGHEVVVVMPEVSWQLGKSLNCTVKTYSTSYTLEDLDREFMDFADAQWKAQV
                     RSLFSLFLSSSNGFFNLFFSHCRSLFNDRRLVEYLKESSFDAVFLDPFDACALIVAKY
                     FSLPSVVFARGIGCHYLEEGAQCPAPLSYVPRILLGFSDAMAFKERVRNHILHLEEHL
                     FCQYFSKNALEIASEILQTPVTAYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT      390 a    371 c    390 g    442 t
ORIGIN
```

```
  Query Match           64.4%;  Score 1493.8;  DB 9;  Length 1593;
  Best Local Similarity  96.1%;  Pred. No. 0;
  Matches 1531;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy      19 atggcttgcacagggtggaccagccccttcctctatgtgtgtctgctgctgacctgt 78
           |||||| |||||||||||||||||||| |||| |||||||| | ||||||||||||||||
Db       1 ATGGCTCGCACAGGGTGGACCAGCCCCATTCCCCTATGTGTTTCTCTGCTGCTGACCTGT 60

Qy      79 ggctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcacc 138
           |||||||| |||||||||||||| |||||||||||||||||||||||| ||||||||||||
Db      61 GGCTTTGCTGAGGCAGGGAAGCTGCTGGTAGTGCCCATGGATGGGAGTCACTGGTTCACC 120

Qy     139 atgaggtcggtggtggagaaactcattctcaggggggcatgaggtggttgtagtcatgcca 198
           |||  |||||||||||||||||| || ||||||||||||||||||||||||||||||||||
Db     121 ATGCAGTCGGTGGTGGAGAAACTTATCCTCAGGGGGCATGAGGTGGTTGTAGTCATGCCA 180

Qy     199 gaggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttca 258
           |||||||||||||||||||| |||||||||||||||||||||||||||| ||||| |||
Db     181 GAGGTGAGTTGGCAACTGGGAAAATCACTGAATTGCACAGTGAAGACTTACTCAACCTCA 240

Qy     259 tatacccctggaggatctggaccgggagttcaaggcttttgcccatgctcaatggaaagca 318
           || || |||||||||||||||||||| |||| || ||| ||| ||||||||||||||||||
Db     241 TACACTCTGGAGGATCTGGACCGGGAATTCATGGATTTCGCCGATGCTCAATGGAAAGCA 300

Qy     319 caagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttattt 378
           |||||||||||| | | ||||||||| || |||||| |||||   ||||| ||||||||
Db     301 CAAGTACGAAGTTTGTTTTCTCTATTTCTGAGTTCATCCAATGGTTTTTTTAACTTATTT 360

Qy     379 ttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagagt 438
           |||||  |||||||||||||||||||| |||  || ||||||||||||||||||||||||||
Db     361 TTTTCGCATTGCAGGAGTTTGTTTAATGACCGAAGATTAGTAGAATACTTAAAGGAGAGT 420

Qy     439 tcttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaatat 498
           |||||||||||| |||||||| |||||||||||  |||||  ||||||||||||||||||
Db     421 TCTTTTGATGCGGTGTTTCTTGATCCTTTTGATGCCTGTGCGTTAATTGTTGCCAAATAT 480

Qy     499 ttctccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaaggt 558
           |||||||||||||| |||||||||||||||||||||  ||||||||||||||||||||||
Db     481 TTCTCCCTCCCCTCTGTGGTCTTCGCCAGGGGAATAGGTTGCCACTATCTTGAAGAAGGT 540

Qy     559 gcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgcc 618
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTCTTAGGGTTCTCAGATGCC 600

Qy     619 atgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatgc 678
           ||| |||| |||||||||||||||||||||||||| ||||||||||||||||||||| |||
Db     601 ATGGCTTTCAAGGAGAGAGTACGGAACCACATCTTGCACTTGGAGGAACATTTATTTTGC 660

Qy     679 caccgttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttacg 738
           ||   |||.| ||||||||||||||||||||||||||||||||||||||||||||| ||
Db     661 CAGTATTTTTCCAAAAATGCCCTAGAAATAGCCTCTGAAATTCTCCAAACACCTGTCACA 720

Qy     739 gagtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttggac 798
           |  ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     721 GCATATGATCTCTACAGCCACACATCAATTTGGTTGTTGCGAACAGACTTTGTTTTGGAC 780

Qy     799 tatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggga 858
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     781 TATCCCAAACCCGTGATGCCCAATATGATCTTCATTGGTGGTATCAACTGCCATCAGGGA 840

Qy     859 aagccgttgcctatggaatttgaagcctacattaatgcttctggagaacatggaattgtg 918
           |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 AAGCCATTGCCTATGGAATTTGAAGCCTACATTAATGCTTCTGGAGAACATGGAATTGTG 900

Qy     919 gttttctctttgggatcaatggtctcagaaattccagagaagaaagctatggcaattgct 978
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATGGCAATTGCT 960

Qy     979 gatgctttcggcaaaatccctcagacagtcctgtggcggtacactggaacccgaccatcg 1038
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GATGCTTTCGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGGAACCCGACCATCG 1020

Qy    1039 aatcttgccaacaacacgatacttgttaagtggctaccccaaaacgatctgcttggtcac 1098
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 AATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACGATCTGCTTGGTCAC 1080

Qy    1099 ccgatgacccgtgcctttatcacccatgctggttcccatggtgtttatgaaagcatatgc 1158
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAGCATATGC 1140

Qy    1159 aatggcgttcccatggtgatgatgcccttgtttggtgatcagatggacaatgcaaagcgc 1218
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 AATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAATGCAAAGCGC 1200

Qy    1219 atggagactaagggagctggagtgaccctgaatgttctggaaatgacttctgaagattta 1278
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGAAGATTTA 1260

Qy    1279 gaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctc 1338
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 GAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAACATCATGCGCCTC 1320

Qy    1339 tccagccttcacaaggaccgcccggtggagccgctggacctggccgtgttctgggtggag 1398
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 TCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAG 1380

Qy    1399 tttgtgatgaggcacaagggcgcgccacacctgcgccccgcagcccacgacctcacctgg 1458
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCCACGACCTCACCTGG 1440
```





```
RESULT  15
AF030310
LOCUS       AF030310               1593 bp    mRNA    linear   PRI 26-OCT-2000
DEFINITION  Homo sapiens UDP-glucuronosyltransferase 1A8 (UGT1A8) mRNA,
            complete cds.
ACCESSION   AF030310
VERSION     AF030310.1  GI:2613043
```

```
Db    1256 CCTTGTTTGGTGATCAGATGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGA 1315

Qy    1244 ccctgaatgttctggaaatgacttctgaagatttagaaaatgctctaaaagcagtcatca 1303
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1316 CCCTGAATGTTCTGGAAATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCA 1375

Qy    1304 atgccaaaagttacaaggagaacatcatgcgcctctccagccttcacaaggaccgcccgg 1363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1376 ATGACAAAAGTTACAAGGAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGG 1435

Qy    1364 tggagccgctggacctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgc 1423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1436 TGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGC 1495

Qy    1424 cacacctgcgccccgcagcccacgacctcacctggtaccagtaccattccttggacgtga 1483
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1496 CACACCTGCGCCCCGCAGCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGA 1555

Qy    1484 ttggtttcctcttggccgtcgtgctgacagtggccttcatcacctttaaatgttgtgctt 1543
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    1556 TTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTCCTT 1615

Qy    1544 atggctaccggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaaga 1603
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1616 ATGGCTACCCGAAATGCTTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGA 1675

Qy    1604 cccattgagaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttcca 1663
           |||||||||| |||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1676 CCCATTGAGAGGTGGGTGGGAAATAAGGTAAAATTTTGAACCA-TCCCTAGTCATTTCCA 1734

Qy    1664 aacttgaaaacagaatcagtgttaaattcattttattcttattaaggaaatactttgcat 1723
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1735 AACTTGAAAACAGAATCAGTGTTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCAT 1794

Qy    1724 aaattaatcagccccagagtg-ctttaaaaaattctcttaaataaaaataatagactcgc 1782
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    1795 AAATTAATCAGCCCCAGAGTGTCTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGC 1854

Qy    1783 tagtcagtaaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatga 1842
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1855 TAGTCAGTAAAGATATTTGAATATGTATCGTGCCCCCTCCGGAGTCTTTGATCAGGATGA 1914

Qy    1843 catgtgccatttttcagaggacgtgcagacaggctggcattctagattacttttcttact 1902
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1915 CATGTGCCATTTTTCAGAGGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACT 1974

Qy    1903 ctgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctact 1962
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1975 CTGAAACATGGCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCACGGCTGCCCCTACT 2034

Qy    1963 gcaaatggcagttttaatcttatcttttggcttctgcagatggttgcaattgatccttaa 2022
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    2035 GCAAATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTTGC-ATTGATCCTTAA 2093

Qy    2023 ccaataatggtcagtcctcatctctgtcctg-cttcataggtgccaccttgtgtgtttaa 2081
           |||||||||||||||||||||||||||||| |||||||||||||||||||| ||||||||||
Db    2094 CCAATAATGGTCAGTCCTCATCTCTGTCCTGACTTCATAGGTGCCACCTTGTG-GTTTAA 2152

Qy    2082 agaagggaagctttgtacctttagagtgtaggtgaaatgaatgaatggcttggagtgcac 2141
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2153 AGAAGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCAC 2212

Qy    2142 tgagaacagcatatgatttcttgctttggggaaaaagaatgatgctatgaaattggtggg 2201
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2213 TGAGAACAGCATATGATTTCTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGG 2272

Qy    2202 tggtgtatttgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaacc 2261
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2273 TGGTGTATTTGAGAAGATAATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACC 2332

Qy    2262 caaaatacagctatgaagtgctgggcaagtttac-ttttttctgatgtttcctacaact 2320
           |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    2333 CAAAATACAGCTATGAAGTGCTGGGCAAGTTTACTTTTTTTTCTGATGTTTCCTACAACT 2392
```

```
RESULT  14
HUMUGT1FA
LOCUS       HUMUGT1FA               2422 bp    mRNA    linear   PRI 14-JAN-1995
DEFINITION  Homo sapiens phenol UDP-glucuronosyltransferase (UDPGT) mRNA,
            complete cds.
ACCESSION   J04093
VERSION     J04093.1  GI:624724
KEYWORDS    UDP-glucuronosyltransferase; phenol.
SOURCE      Homo sapiens kidney cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2422)
  AUTHORS   Harding,D., Fournel-Gigleux,S., Jackson,M.R. and Burchell,B.
  TITLE     Cloning and substrate specificity of a human phenol
            UDP-glucuronosyltransferase expressed in COS-7 cells
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85 (22), 8381-8385 (1988)
  MEDLINE   89042187
COMMENT     On Jan 18, 1995 this sequence version replaced gi:340517.
FEATURES             Location/Qualifiers
     source          1. .2422
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="kidney"
     gene            1. .2422
                     /gene="UDPGT"
     CDS             88. .1683
                     /gene="UDPGT"
                     /EC_number="2.4.1.17"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase"
                     /protein_id="AAA61251.1"
                     /db_xref="GI:624725"
                     /translation="MACLLRSFQRISAGVFFLALWGMVVGDKLLVVPQDGSHWLSMKD
                     IVEVLSDRGHEIVVVVPEVNLLLKEYKYYTRKIYPVPYDQEELKNRYQSFGNNHFAER
                     SFLTAPQTEYRNNMIVIGLYFINCQSLLQDRDTLNFFKESKFDALFTDPALPCGVILA
                     EYLGLPSVYLFRGFPCSLEHTFSRSPDPVSYIPRCYTKFSDHMTFSQRVANFLVNLLE
                     PYLFYCLFSKYEKLASAVLKRDVDIITLSEVSVWLLRYDFVLEYPRPVMPNMVFIGGI
                     NCKKRKDLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKNPQTVLWR
                     YTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLF
                     GDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPV
                     EPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCC
                     PYGYPKCLGKKGRVKKAHKSKTH"
     polyA_signal    2396. .2401
                     /gene="UDPGT"
     polyA_signal    2404. .2409
                     /gene="UDPGT"
     polyA_site      2422
                     /gene="UDPGT"
BASE COUNT      653 a    516 c    555 g    698 t
ORIGIN
```

```
  Query Match             66.3%;  Score 1539.2;  DB 9;  Length 2422;
  Best Local Similarity   82.0%;  Pred. No. 0;
  Matches 1870;  Conservative    0;  Mismatches  398;  Indels   12;  Gaps    8;

Qy      47 ttcctctatgtgtgtctgctgctgacctgtggctttgccgaggcagggaagctactgg 106
           || ||  | | || |   |  |       || ||| | |  |  |  |  ||||| ||||
Db     119 TTTCTGCAGGGGTTTTCTTCTTAGCTCTTTGGGGCATGGTTGTAGGTGACAAGCTGCTGG 178

Qy     107 tagtgcccatggatgggagccactggttcaccatgaggtcggtggtggagaaactcattc 166
           | || ||   ||| || |||||||||| | |  |||| |    | || |||   |||| |
Db     179 TGGTCCCTCAGGACGGAAGCCACTGGCTTAGTATGAAGGATATAGTTGAGGTTCTCAGTG 238

Qy     167 tcaggggcatgaggtggttgtagtcatgccagaggtgagttggcaactgggaagatcac 226
           | |||| |||||| | || || ||  |||| || || | || ||   ||  |  ||
Db     239 ACCGGGGTCATGAGATTGTAGTGGTGGTGCCTGAAGTTAATTTGCTTTTGAAAGAATACA 298

Qy     227 tgaattgcacagtgaagacttattcaacttcatataccctggaggatctggaccgggagt 286
             | | ||||   || |  ||| ||    | |||  ||  || || ||| |       |
```

```
Qy   1805 atgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggac 1864
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1811 ATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGAC 1870

Qy   1865 gtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttggga 1924
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1871 GTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGA 1930

Qy   1925 gtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatctta 1984
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1931 GTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTA 1990

Qy   1985 tcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatc 2044
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1991 TCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCTCATC 2050

Qy   2045 tctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtaccttta 2104
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2051 TCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTTA 2110

Qy   2105 gagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttcttg 2164
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2111 GAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTTCTTG 2170

Qy   2165 ctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatca 2224
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2171 CTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATAATCA 2230

Qy   2225 ttgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgctg 2284
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2231 TTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGTGCTG 2290

Qy   2285 ggcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||||
Db   2291 GGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 2326
```

```
RESULT  13
AX336711
LOCUS       AX336711                2422 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 7220 from Patent WO0194629.
ACCESSION   AX336711
VERSION     AX336711.1  GI:18127430
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 7220 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1. .2422
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      653 a    516 c    555 g    698 t
ORIGIN
```

```
  Query Match             66.3%;  Score 1539.2;  DB 6;  Length 2422;
  Best Local Similarity   82.0%;  Pred. No. 0;
  Matches 1870;  Conservative    0;  Mismatches  398;  Indels   12;  Gaps    8;

Qy      47 ttcctctatgtgtgtgtctgctgctgacctgtggctttgccgaggcagggaagctactgg 106
           || ||  | | || |   |  |       || ||| | |  |  |  | ||||| ||||
Db     119 TTTCTGCAGGGGTTTTCTTCTTAGCTCTTTGGGGCATGGTTGTAGGTGACAAGCTGCTGG 178

Qy     107 tagtgcccatggatgggagccactggttcaccatgaggtcggtggtggagaaactcattc 166
```

```
                | || ||   ||| || ||||||||| | |  |||| |    | || |||   |||| |
Db     179 TGGTCCCTCAGGACGGAAGCCACTGGCTTAGTATGAAGGATATAGTTGAGGTTCTCAGTG 238

Qy     167 tcagggggcatgaggtggttgtagtcatgccagaggtgagttggcaactgggaagatcac 226
           | |||| |||||| | || || ||  |||| || || | || ||   ||  | ||
Db     239 ACCGGGGTCATGAGATTGTAGTGGTGGTGCCTGAAGTTAATTTGCTTTTGAAAGAATACA 298

Qy     227 tgaattgcacagtgaagacttattcaacttcatataccctggaggatctggaccgggagt 286
             | | ||||   || |  ||| ||    | |||  ||  || || ||| |        |
Db     299 AATACTACACAAGAAAAATCTATCCAGTGCCGTATGACCAAGAAGAGCTGAAGAACCGTT 358

Qy     287 tcaaggcttttgcccatgctcaatggaaagcacaagtacgaagtatatattctctattaa 346
           | |  | ||||   |  ||| |     | | | |     |   | |||   |
Db     359 ACCAATCATTTGGAAACAATCACTTTGCTGAGCGATCATTCCTAACTGCTCCTCAGACAG 418

Qy     347 tgggttcatacaatgaca---tttttgacttatttttttcaaattgcaggagtttgttta 403
           |      | ||   |   || ||| | | |  ||    || ||| ||| |  |
Db     419 AGTACAGGAATAACATGATTGTTATTGGCCTGTACTTCATCAACTGCCAGAGCCTCCTGC 478

Qy     404 aagacaaaaaattagtagaatacttaaaggagagttcttttgatgcagtgtttctcgatc 463
           | ||||   |     | | | ||| ||||||||    ||||||||  | ||    || |
Db     479 AGGACAGGGACACCCTGAACTTCTTTAAGGAGAGCAAGTTTGATGCTCTTTTCACAGACC 538

Qy     464 cttttgataactgtggcttaattgttgccaaatatttctccctcccctccgtggtcttcg 523
           |         ||||||  | ||  | ||  | |||||   ||| || || ||| |  ||
Db     539 CAGCCTTACCCTGTGGGGTGATCCTGGCTGAGTATTTGGGCCTACCATCTGTGTACCTCT 598

Qy     524 ccaggggaatactttgccactatcttgaagaaggtgcacagtgccctgctcctctttcct 583
           ||||||  | |  ||   |      |   |          |||  |  ||| | ||||
Db     599 TCAGGGGTTTTCCGTGTTCCCTGGAGCATACATTCAGCAGAAGCCCAGACCCTGTGTCCT 658

Qy     584 atgtccccagaattctcttagggttctcagatgccatgactttcaaggagagagtacgga 643
           | | ||||      |  | ||| |||||  |||||||||     |  ||||    |
Db     659 ACATTCCCAGGTGCTACACAAAGTTTTCAGACCACATGACTTTTTCCCAACGAGTGGCCA 718

Qy     644 accacatcatgcacttggaggaacatttattatgccaccgtttttttcaaaaatgccctag 703
           ||  | |  |  | |||  ||| |  |   |||   |  || | ||   |||
Db     719 ACTTCCTTGTTAATTTGTTGGAGCCCTATCTATTTTATTGTCTGTTTTCAAAGTATGAGA 778

Qy     704 aaatagcctctgaaattctccaaacacctgttacggagtatgatctctacagccacacat 763
           || | || || |   | ||| | | |  |||   |||          ||     |   |
Db     779 AACTCGCATCAGCTGTCCTCAAGAGAGATGT---GGATATAATCACCTTATCAGAGGTCT 835

Qy     764 caatttggttgttgcgaacggactttgttttggactatcccaaacccgtgatgcccaaca 823
           |  ||||| ||||  ||   |||||||  | || ||||| |  || || ||||||||||
Db     836 CTGTTTGGCTGTTAAGATATGACTTTGTGCTTGAATATCCTAGGCCGGTCATGCCCAACA 895

Qy     824 tgatcttcattggtggtatcaactgccatcagggaaagccgttgcctatggaatttgaag 883
           || |||||||||| |||||||||||  |  || | ||   ||| ||  ||||||||||
Db     896 TGGTCTTCATTGGAGGTATCAACTGTAAGAAGAGGAAAGACTTGTCTCAGGAATTTGAAG 955

Qy     884 cctacattaatgcttctggagaacatggaattgtggttttctctttgggatcaatggtct 943
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     956 CCTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCT 1015

Qy     944 cagaaattccagagaagaaagctatggcaattgctgatgctttgggcaaaatccctcaga 1003
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    1016 CAGAAATTCCAGAGAAGAAAGCTATGGCAATTGCTGATGCTTTGGGCAAAAACCCTCAGA 1075

Qy    1004 cagtcctgtggcggtacactggaacccgaccatcgaatcttgcgaacaacacgatacttg 1063
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1076 CAGTCCTGTGGCGGTACACTGGAACCCGACCATCGAATCTTGCGAACAACACGATACTTG 1135

Qy    1064 ttaagtggctaccccaaaacgatctgcttggtcacccgatgaccgtgccttttatcaccc 1123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1136 TTAAGTGGCTACCCCAAAACGATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCC 1195

Qy    1124 atgctggttcccatggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgc 1183
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1196 ATGCTGGTTCCCATGGTGTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGC 1255

Qy    1184 ccttgtttggtgatcagatggacaatgcaaagcgcatggagactaagggagctggagtga 1243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
                     /protein_id="AAA63195.1"
                     /db_xref="GI:184473"
                     /db_xref="GDB:G00-120-007"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQLQQRGHEIVVLAPDASLYIRDGAFYTLKTYPVPFQREDVKESFVSLGHNVFEN
                     DSFLQRVIKTYKKIKKDSAMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCSPIV
                     AQYLSLPTVFFLHALPCSLEFEATQCPNPFSYVPRPLSSHSDHMTFLQRVKNMLIAFS
                     QNFLCDVVYSPYATLASEFLQREVTVQDLLSSASVWLFRSDFVKDYPRPIMPNMVFVG
                     GINCLHQNPLSQEPEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVL
                     WRYTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMP
                     LFGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDR
                     PVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFK
                     CCAYGYRKCLGKKGRVKKAHKSKTH"
     sig_peptide     43. .75
                     /gene="UGT1"
                     /note="G00-120-007"
     mat_peptide     76. .1614
                     /gene="UGT1"
                     /standard_name="bilirubin UDP-glucuronosyltransferase
                     isozyme 1"
                     /EC_number="2.4.1.17"
                     /note="G00-120-007"
                     /product="UDP-glucuronosyltransferase 1"
BASE COUNT      601 a    540 c    557 g    653 t
ORIGIN
```

Query Match 69.4%; Score 1610.8; DB 9; Length 2351;
Best Local Similarity 82.7%; Pred. No. 0;
Matches 1866; Conservative 0; Mismatches 387; Indels 3; Gaps 2;

```
Qy      68 tgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatgggagcc 127
           |||||  |   |   |  || | || |||||| |||||  |  | ||  |||||||  ||||
Db      71 TGCTGGGCCCAGTGGTGTCCCATGCTGGGAAGATACTGTTGATCCCAGTGGATGGCAGCC 130

Qy     128 actggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgaggtggttg 187
           ||||| | | ||||     ||   |  || | ||    |  ||||| |||||  | ||||
Db     131 ACTGGCTGAGCATGCTTGGGGCCATCCAGCAGCTGCAGCAGAGGGGACATGAAATAGTTG 190

Qy     188 tagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaagactt 247
           |  |    || || |     | | |  |  ||      |    || |||  |||||||  |
Db     191 TCCTAGCACCTGACGCCTCGTTGTACATCAGAGACGGAGCATTTTACACCTTGAAGACGT 250

Qy     248 attcaacttcatataccctggaggatctggaccgggagtt--caaggcttttgcccatgc 305
           |  |     |||       |||||||  || |   |   ||    || ||   ||  |
Db     251 ACCCTGTGCCATTCCAAAGGGAGGATGTGAAAGAGTCTTTTGTTAGTCTCGGGCATAATG 310

Qy     306 tcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgacat 365
           |   ||  ||  |    | |         | |  || |   ||      | | ||  |
Db     311 TTTTTGAGAATGATTCTTTCCTGCAGCGTGTGATCAAAACATACAAGAAAATAAAAAAGG 370

Qy     366 ttttgact-tatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaat 424
             |  || |  |||| ||    ||       ||  |  |  ||||  |  |  | |  |
Db     371 ACTCTGCTATGCTTTTGTCTGGCTGTTCCCACTTACTGCACAACAAGGAGCTCATGGCCT 430

Qy     425 acttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggcttaa 484
           | |    || ||    |||||||   || |    || |||||    |   ||  ||   |
Db     431 CCCTGGCAGAAAGCAGCTTTGATGTCATGCTGACGGACCCTTTCCTTCCTTGCAGCCCCA 490

Qy     485 ttgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgccact 544
           | || ||| | ||  | || || ||| | ||  |||||      | | | |  |||  |
Db     491 TCGTGGCCCAGTACCTGTCTCTGCCCACTGTATTCTTCTTGCATGCACTGCCATGCAGCC 550

Qy     545 atcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttag 604
               || | | |||||||||    ||  | ||||| || |||||   |||||
Db     551 TGGAATTTGAGGCTACCCAGTGCCCCAACCCATTCTCCTACGTGCCCAGGCCTCTCTCCT 610

Qy     605 ggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggagg 664
           |||||  ||||||  ||| | || | ||   ||||    ||||   |||
Db     611 CTCATTCAGATCACATGACCTTCCTGCAGCGGGTGAAGAACATGCTCATTGCCTTTTCAC 670

Qy     665 aacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaattctcc 724
           |  | ||  | ||| ||    |||              |    | ||||| ||| | || |
```

```
Db     671 AGAACTTTCTGTGCGACGTGGTTTATTCCCCGTATGCAACCCTTGCCTCAGAATTCCTTC 730

Qy     725 aaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacgg 784
           | | |   || || |   | || || |  |||    ||||  | ||| ||||  |||  |
Db     731 AGAGAGAGGTGACTGTCCAGGACCTATTGAGCTCTGCATCTGTCTGGCTGTTTAGAAGTG 790

Qy     785 actttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatca 844
           |||||||   ||| || || |  ||| | ||||||| ||| | ||  |||||| ||||
Db     791 ACTTTGTGAAGGATTACCCTAGGCCCATCATGCCCAATATGGTTTTTGTTGGTGGAATCA 850

Qy     845 actgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctggag 904
           ||||||  |||   ||| ||  |  |  ||||||||||||||||||||||||||||||||
Db     851 ACTGCCTTCACCAAAATCCACTATCCCAGGAATTTGAAGCCTACATTAATGCTTCTGGAG 910

Qy     905 aacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaagaaag 964
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     911 AACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAG 970

Qy     965 ctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactg 1024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     971 CTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTG 1030

Qy     1025 gaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacg 1084
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1031 GAACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACG 1090

Qy     1085 atctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgttt 1144
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1091 ATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTT 1150

Qy     1145 atgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatgg 1204
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1151 ATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGG 1210

Qy     1205 acaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatga 1264
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1211 ACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGA 1270

Qy     1265 cttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggaga 1324
            |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1271 CTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGA 1330

Qy     1325 acatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccg 1384
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1331 ACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCG 1390

Qy     1385 tgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagccc 1444
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1391 TGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCC 1450

Qy     1445 acgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcg 1504
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1451 ACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCG 1510

Qy     1505 tgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttgg 1564
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1511 TGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGG 1570

Qy     1565 ggaaaaaacggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtggga 1624
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1571 GGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGA 1630

Qy     1625 aataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtg 1684
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1631 AATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTG 1690

Qy     1685 ttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtg 1744
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1691 TTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTG 1750

Qy     1745 ctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaat 1804
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1751 CTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAAT 1810
```

```
Db    726 AACTCGCATCAGCTGTCCTCAAGAGAGATGTGGATATAATCACCTTATATCAGAAGGTCT 785

Qy    764 caatttggttgttgcgaacggactttgttttggactatcccaaaccccgtgatgcccaaca 823
          |  ||||| |||| ||   ||||||||  | || ||||| |  || || ||||||||||
Db    786 CTGTTTGGCTGTTAAGATATGACTTTGTGCTTGAATATCCTAGGCCGGTCATGCCCAACA 845

Qy    824 tgatcttcattggtggtatcaactgccatcagggaaagccgttgcctatggaatttgaag 883
          || |||||||||| ||||||||||   |  || | ||    ||| ||  ||||||||||
Db    846 TGGTCTTCATTGGAGGTATCAACTGTAAGAAGAGGAAAGACTTGTCTCAGGAATTTGAAG 905

Qy    884 cctacattaatgcttctggagaacatggaattgtggttttctctttgggatcaatggtct 943
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    906 CCTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCT 965

Qy    944 cagaaattccagagaagaaagctatggcaattgctgatgctttgggcaaaatccctcaga 1003
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    966 CAGAAATTCCAGAGAAGAAAGCTATGGCAACTGCTGATGCTTTGGGCAAAATCCCTCAGA 1025

Qy   1004 cagtcctgtggcggtacactggaacccgaccatcgaatcttgcgaacaacacgatacttg 1063
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1026 CAGTCCTGTGGCGGTACACTGGAACCCGACCATCGAATCTTGCGAACAACACGATACTTG 1085

Qy   1064 ttaagtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcctttatcaccc 1123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1086 TTAAGTGGCTACCCCAAAACGATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCC 1145

Qy   1124 atgctggttcccatggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgc 1183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1146 ATGCTGGTTCCCATGGTGTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGC 1205

Qy   1184 ccttgtttggtgatcagatggacaatgcaaagcgcatggagactaagggagctggagtga 1243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1206 CCTTGTTTGGTGATCAGATGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGA 1265

Qy   1244 ccctgaatgttctggaaatgacttctgaagatttagaaaatgctctaaaagcagtcatca 1303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1266 CCCTGAATGTTCTGGAAATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCA 1325

Qy   1304 atgacaaaagttacaaggagaacatcatgcgcctctccagccttcacaaggaccgcccgg 1363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1326 ATGACAAAAGTTACAAGGAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGG 1385

Qy   1364 tggagccgctggacctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgc 1423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1386 TGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGC 1445

Qy   1424 cacacctgcgccccgcagcccacgacctcacctggtaccagtaccattccttggacgtga 1483
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1446 CACACCTGCGCCCCGCAGCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGA 1505

Qy   1484 ttggtttcctcttggccgtcgtgctgacagtggccttcatcacctttaaatgttgtgctt 1543
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1506 TTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTT 1565

Qy   1544 atggctaccggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaaga 1603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1566 ATGGCTACCGGAAATGCTTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGA 1625

Qy   1604 cccattgagaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttcca 1663
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1626 CCCATTGAGAAGTGGGTGGGAAATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCA 1685

Qy   1664 aacttgaaaacagaatcagtgttaaattcattttattcttattaaggaaatactttgcat 1723
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1686 AACTTGAAAACAGAATCAGTGTTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCAT 1745

Qy   1724 aaattaatcagccccagagtgctttaaaaaattctcttaaataaaaataatagactcgct 1783
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1746 AAATTAATCAGCCCCAGAGTGCTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCT 1805

Qy   1784 agtcagtaaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgac 1843
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   1806 AGTCAGTAAAGATATTTGAATATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGAC 1865

Qy   1844 atgtgccatttttcagaggacgtgcagacaggctggcattctagattacttttcttactc 1903
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1866 ATGTGCCATTTTTCAGAGGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTC 1925

Qy   1904 tgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctactg 1963
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1926 TGAAACATGGCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTG 1985

Qy   1964 caaatggcagttttaatcttatcttttggcttctgcagatggttgcaattgatccttaac 2023
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1986 CAAATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAAC 2045

Qy   2024 caataatggtcagtcctcatctctgtcctgcttcataggtgccaccttgtgtgtttaaag 2083
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2046 CAATAATGGTCAGTCCTCATCTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAG 2105

Qy   2084 aagggaagctttgtacctttagagtgtaggtgaaatgaatgaatggcttggagtgcactg 2143
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2106 AAGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTG 2165

Qy   2144 agaacagcatatgatttcttgctttggggaaaaagaatgatgctatgaaattggtgggtg 2203
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2166 AGAACAGCATATGATTTCTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTG 2225

Qy   2204 gtgtatttgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaaccca 2263
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2226 GTGTATTTGAGAAGATAATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCA 2285

Qy   2264 aaatacagctatgaagtgctgggcaagtttacttttttttctgatgtttcctacaact 2320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2286 AAATACAGCTATGAAGTGCTGGGCAAGTTTACTTTTTTTTCTGATGTTTCCTACAACT 2342
```

```
RESULT  12
HUMHUGBR1
LOCUS       HUMHUGBR1               2351 bp    mRNA    linear   PRI 07-MAR-1995
DEFINITION  Human bilirubin UDP-glucuronosyltransferase isozyme 1 mRNA,
            complete cds.
ACCESSION   M57899
VERSION     M57899.1  GI:184472
KEYWORDS    UDP-glucuronosyltransferase; bilirubin UDP-glucuronosyltransferase.
SOURCE      Human adult female liver, cDNA to mRNA, clones Z6, Z11 and Z6MB2.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2351)
  AUTHORS   Ritter,J.K., Crawford,J.M. and Owens,I.S.
  TITLE     Cloning of two human liver bilirubin UDP-glucuronosyltransferase
            cDNAs with expression in COS-1 cells
  JOURNAL   J. Biol. Chem. 266 (2), 1043-1047 (1991)
  MEDLINE   91093210
FEATURES             Location/Qualifiers
     source          1. .2351
                     /organism="Homo sapiens"
                     /isolate="AK"
                     /db_xref="taxon:9606"
                     /map="2"
                     /clone="Z6"
                     /sex="female"
                     /tissue_type="liver"
                     /dev_stage="adult"
                     /tissue_lib="lambda-ZAP"
     gene            1. .2351
                     /gene="UGT1"
     CDS             16. .1617
                     /gene="UGT1"
                     /standard_name="bilirubin UDP-glucuronosyltransferase
                     isozyme 1"
                     /EC_number="2.4.1.17"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1"
```

```
Db   1901 GGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGAGTGCGGGATTCAA 1960

Qy   1938 aggtggtcccaccgctgcccctactgcaaatggcagttttaatcttatctttggcttct 1997
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1961 AGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTATCTTTTGGCTTCT 2020

Qy   1998 gcagatggttgcaattgatccttaaccaataatggtcagtcctcatctctgtcctgcttc 2057
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2021 GCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCTCATCTCTGTCCTGCTTC 2080

Qy   2058 ataggtgccaccttgtgtgtttaaagaagggaagctttgtacctttagagtgtaggtgaa 2117
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2081 ATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAA 2140

Qy   2118 atgaatgaatggcttggagtgcactgagaacagcatatgatttcttgctttggggaaaaa 2177
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2141 ATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTTCTTGCTTTGGGGAAAAA 2200

Qy   2178 gaatgatgctatgaaattggtgggtggtgtatttgagaagataatcattgcttatgtcaa 2237
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2201 GAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATAATCATTGCTTATGTCAA 2260

Qy   2238 atggagctgaatttgataaaaacccaaaatacagctatgaagtgctgggcaagtttactt 2297
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2261 ATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGTGCTGGGCAAGTTTACTT 2320

Qy   2298 tttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||
Db   2321 TTTTTCTGATGTTTCCTACAACT 2343


RESULT  11
BC019861
LOCUS       BC019861               2388 bp    mRNA    linear   PRI 22-JAN-2002
DEFINITION  Homo sapiens, clone MGC:29860 IMAGE:4454322, mRNA, complete cds.
ACCESSION   BC019861
VERSION     BC019861.1  GI:18043700
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2388)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 42 Row: c Column: 8
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4507814.
FEATURES             Location/Qualifiers
     source          1. .2388
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:29860 IMAGE:4454322"
                     /tissue_type="Kidney, hypernephroma"
                     /clone_lib="NIH_MGC_89"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             35. .1633
                     /codon_start=1
                     /product="Unknown (protein for MGC:29860)"
                     /protein_id="AAH19861.1"
                     /db_xref="GI:18043701"
                     /translation="MACLLRSFQRISAGVFFLALWGMVVGDKLLVVPQDGSHWLSMKD
                     IVEVLSDRGHEIVVVVPEVNLLLKESKYYTRKIYPVPYDQEELKNRYQSFGNNHFAER
                     SFLTAPQTEYRNNMIVIGLYFINCQSLLQDRDTLNFFKESKFDALFTDPALPCGVILA
                     EYLGLPSVYLFRGFPCSLEHTFSRSPDPVSYIPRCYTKFSDHMTFSQRVANFLVNLLE
                     PYLFYCLFSKYEELASAVLKRDVDIITLYQKVSVWLLRYDFVLEYPRPVMPNMVFIGG
                     INCKKRKDLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMATADALGKIPQTVLW
                     RYTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPL
                     FGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRP
                     VEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKC
                     CAYGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT      648 a    503 c    546 g    691 t
ORIGIN


  Query Match             70.2%;  Score 1629;  DB 9;  Length 2388;
  Best Local Similarity   82.5%;  Pred. No. 0;
  Matches 1879;  Conservative    0;  Mismatches  395;  Indels    3;  Gaps    1;

Qy     47 ttcctctatgtgtgtctgctgctgacctgtggctttgccgaggcagggaagctactgg 106
          || ||  | | || |   |  |       || ||| | |  |  |  |  ||||| ||||
Db     66 TTTCTGCAGGGGTTTTCTTCTTAGCACTTTGGGGCATGGTTGTAGGTGACAAGCTGCTGG 125

Qy    107 tagtgcccatggatgggagccactggttcaccatgaggtcggtggtggagaaactcattc 166
          | || ||   ||| || ||||||||| | |  |||| |    | || |||   |||| |
Db    126 TGGTCCCTCAGGACGGAAGCCACTGGCTTAGTATGAAGGATATAGTTGAGGTTCTCAGTG 185

Qy    167 tcaggggcatgaggtggttgtagtcatgccagaggtgagttggcaactgggaagatcac 226
          | |||| |||||| | || || ||  |||| || || | || ||   ||  |  |||
Db    186 ACCGGGGTCATGAGATTGTAGTGGTGGTGCCTGAAGTTAATTTGCTTTTGAAAGAATCCA 245

Qy    227 tgaattgcacagtgaagacttattcaacttcatataccctggaggatctggaccgggagt 286
          | | ||||   || |  ||| ||    | |||  ||  || || |||| |       |
Db    246 AATACTACACAAGAAAAATCTATCCAGTGCCGTATGACCAAGAAGAGCTGAAGAACCGTT 305

Qy    287 tcaaggcttttgcccatgctcaatggaaagcacaagtacgaagtatatattctctattaa 346
          | |  | ||||   |   ||| |     | | |  |      |    | |||    |
Db    306 ACCAATCATTTGGAAACAATCACTTTGCTGAGCGATCATTCCTAACTGCTCCTCAGACAG 365

Qy    347 tgggttcatacaatgaca---tttttgacttatttttttcaaattgcaggagtttgttta 403
          |       | ||    |   || ||| | | |  ||    || |||  |||  |  |
Db    366 AGTACAGGAATAACATGATTGTTATTGGCCTGTACTTCATCAACTGCCAGAGCCTCCTGC 425

Qy    404 aagacaaaaaaattagtagaatacttaaaggagagttcttttgatgcagtgtttctcgatc 463
          | ||||   |     |  | | ||| ||||||||    ||||||||  | ||    || |
Db    426 AGGACAGGGACACCCTGAACTTCTTTAAGGAGAGCAAGTTTGATGCTCTTTCACAGACC 485

Qy    464 cttttgataactgtggcttaattgttgccaaatatttctccctccccctccgtggtcttcg 523
          |       ||||||  | ||  | ||  | |||||   ||| || || |||  | ||
Db    486 CAGCCTTACCCTGTGGGGTGATCCTGGCTGAGTATTTGGGCCTACCATCTGTGTACCTCT 545

Qy    524 ccaggggaatactttgccactatcttgaagaaggtgcacagtgccctgctcctctttcct 583
          ||||||  | |  ||   |       |   |          |||| |  ||| | ||||
Db    546 TCAGGGGTTTCCGTGTTCCCTGGAGCATACATTCAGCAGAAGCCCAGACCCTGTGTCCT 605

Qy    584 atgtccccagaattctcttagggttctcagatgccatgactttcaaggagagagtacgga 643
          |  | |||||      |  |  ||| |||||   |||||||||     |  ||||    |
Db    606 ACATTCCCAGGTGCTACACAAAGTTTTCAGACCACATGACTTTTTCCCAACGAGTGGCCA 665

Qy    644 accacatcatgcacttggaggaacatttattatgccaccgtttttttcaaaaatgccctag 703
          ||  | |  |  | |||  ||| |  |   |||   |  || | ||   |||      ||
Db    666 ACTTCCTTGTTAATTTGTTGGAGCCCTATCTATTTTATTGTCTGTTTTCAAAGTATGAAG 725

Qy    704 aaatagcctctgaaattctccaaacacctgttacggagtatgatctctacagccacacat 763
```

```
                 LAKYLSIPAVFFWRYIPCDLDFKGTQCPNPSSYIPKLLTTNSDHMTFLQRVKNMLYPL
                 ALSYICHTFSAPYASLASELFQREVSVVDLVSYASVWLFRGDFVMDYPRPIMPNMVFI
                 GGINCANGKPLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTV
                 LWRYTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMM
                 PLFGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKD
                 RPVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITF
                 KCCAYGYRKCLGKKGRVKKAHKSKTH"
     sig_peptide     63. .95
                     /gene="UGT2B"
                     /note="G00-127-753"
     mat_peptide     96. .1631
                     /gene="UGT2B"
                     /EC_number="2.4.1.17"
                     /note="G00-127-753"
                     /product="UDP-glucuronosyltransferase 2 family polypeptide
                     B"
BASE COUNT      604 a    534 c    573 g    657 t
ORIGIN
```

Query Match 70.5%; Score 1636.6; DB 9; Length 2368;
Best Local Similarity 83.4%; Pred. No. 0;
Matches 1871; Conservative 0; Mismatches 369; Indels 3; Gaps 1;

```
Qy      81 ctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcaccat 140
           ||  || |||   || ||| |  |||| |||||||  ||||| |||| |||| ||| |||
Db     101 CTGGGCTGAGAGTGGAAAGGTGTTGGTGGTGCCCACTGATGGCAGCCCCTGGCTCAGCAT 160

Qy     141 gaggtcggtggtggagaaactcattctcaggggcatgaggtggttgtagtcatgccaga 200
           | ||  ||   ||  | | |||  |  ||| || ||  ||| ||| ||  |||  |||||
Db     161 GCGGGAGGCCTTGCGGGAGCTCCATGCCAGAGGCCACCAGGCGGTGGTCCTCACCCCAGA 220

Qy     201 ggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttcata 260
           |||||  |  |||  |   |  |        || |||  |||   | ||| |   | |||
Db     221 GGTGAATATGCACATCAAAGAAGAGAAATTTTTCACCCTGACAGCCTATGCTGTTCCATG 280

Qy     261 taccctggaggatctggaccgggagttcaaggcttttgcccatgctcaat---ggaaagc 317
           ||||  | ||||  | || || |       ||  |   | || |     |     | ||
Db     281 GACCCAGAAGGAATTTGATCGCGTTACGCTGGGCTACACTCAAGGGTTCTTTGAAACAGA 340

Qy     318 acaagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttatt 377
           |||  | |  |  | |||||||  |  ||||        ||  |  | | |       |
Db     341 ACATCTTCTGAAGAGATATTCTAGAAGTATGGCAATTATGAACAATGTATCTTTGGCCCT 400

Qy     378 tttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagag 437
           |  |     |||  |    |  |  |  |  |     |  |     || | || |  |
Db     401 TCATAGGTGTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGGCACCTGAATGCTAC 460

Qy     438 ttcttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaata 497
           ||| |||||||  || ||    || ||  || |  ||| ||     |  | || || ||
Db     461 TTCCTTTGATGTGGTTTTAACAGACCCCGTTAACCTCTGCGGGGCGGTGCTGGCTAAGTA 520

Qy     498 tttctccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaagg 557
             | ||  | ||  | ||| | ||    |||   || |  ||  |||         | ||
Db     521 CCTGTCGATTCCTGCTGTGTTTTTTTGGAGGTACATTCCATGTGACTTAGACTTTAAGGG 580

Qy     558 tgcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgc 617
             |||||| ||   ||||  |||||| | || |   | ||     |||||
Db     581 CACACAGTGTCCAAATCCTTCCTCCTATATTCCTAAGTTACTAACGACCAATTCAGACCA 640

Qy     618 catgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatg 677
           |||||| |||  | | || ||  ||||    ||   |  |||       |   | ||
Db     641 CATGACATTCCTGCAAAGGGTCAAGAACATGCTCTACCCTCTGGCCCTGTCCTACATTTG 700

Qy     678 ccaccgtttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttac 737
           |||   |||||       |     |    | ||||||||  || | || | |   ||  |
Db     701 CCATACTTTTTCTGCCCCTTATGCAAGTCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTC 760

Qy     738 ggagtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttgga 797
             | |   |||||  |||| |  ||||  | ||| |||| |||  |||||||||  ||||
Db     761 AGTGGTGGATCTTGTCAGCTATGCATCCGTGTGGCTGTTCCGAGGGGACTTTGTGATGGA 820

Qy     798 ctatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggg 857
```

```
                ||| ||||  ||  | |||||||||||| |||||||||| || ||||||||     | ||
Db     821 CTACCCCAGGCCGATCATGCCCAACATGGTCTTCATTGGGGGCATCAACTGTGCCAACGG 880

Qy     858 aaagccgttgcctatggaatttgaagcctacattaatgcttctggagaacatggaattgt 917
           |||||  |  ||  ||||||||||||||||||||||||||||||||||||||||||||||
Db     881 GAAGCCACTATCTCAGGAATTTGAAGCCTACATTAATGCTTCTGGAGAACATGGAATTGT 940

Qy     918 ggttttctctttgggatcaatggtctcagaaattccagagaagaaagctatggcaattgc 977
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     941 GGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATGGCAATTGC 1000

Qy     978 tgatgctttgggcaaaatccctcagacagtcctgtggcggtacactggaacccgaccatc 1037
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1001 TGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGGAACCCGACCATC 1060

Qy    1038 gaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacgatctgcttggtca 1097
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1061 GAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACGATCTGCTTGGTCA 1120

Qy    1098 cccgatgacccgtgcctttatcacccatgctggttcccatggtgtttatgaaagcatatg 1157
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1121 CCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAGCATATG 1180

Qy    1158 caatggcgttcccatggtgatgatgcccttgtttggtgatcagatggacaatgcaaagcg 1217
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1181 CAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAATGCAAAGCG 1240

Qy    1218 catggagactaagggagctggagtgaccctgaatgttctggaaatgacttctgaagattt 1277
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1241 CATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGAAGATTT 1300

Qy    1278 agaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcct 1337
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1301 AGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAACATCATGCGCCT 1360

Qy    1338 ctccagccttcacaaggaccgcccggtggagccgctggacctggccgtgttctgggtgga 1397
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1361 CTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGA 1420

Qy    1398 gtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagcccacgacctcacctg 1457
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1421 GTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCCACGACCTCACCTG 1480

Qy    1458 gtaccagtaccattccttggacgtgattggtttcctcttggccgtcgtgctgacagtggc 1517
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1481 GTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGC 1540

Qy    1518 cttcatcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcg 1577
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1541 CTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGGAAAAAAGGGCG 1600

Qy    1578 agttaagaaagcccacaaatccaagacccattgagaagtgggtgggaaataaggtaaaat 1637
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1601 AGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGAAATAAGGTAAAAT 1660

Qy    1638 tttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgttaaattcatttt 1697
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1661 TTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTGTTAAATTCATTTT 1720

Qy    1698 attcttattaaggaaatactttgcataaattaatcagccccagagtgctttaaaaaattc 1757
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1721 ATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTGCTTTAAAAAATTC 1780

Qy    1758 tcttaaataaaaataatagactcgctagtcagtaaagatatttgaatatgtatcgtgccc 1817
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1781 TCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAATATGTATCGTGCCC 1840

Qy    1818 cctccggtgtctttgatcaggatgacatgtgccatttttcagaggacgtgcagacaggct 1877
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1841 CCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGACGTGCAGACAGGCT 1900

Qy    1878 ggcattctagattacttttcttactctgaaacatggcctgtttgggagtgcgggattcaa 1937
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy    846 ctgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctggaga 905
          ||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    481 CTGTCATCAGGGAAAGCCATTGCCTATGGAATTTGAAGCCTACATTAATGCTTCTGGAGA 540

Qy    906 acatggaattgtggtttttctctttgggatcaatggtctcagaaattccagagaagaaagc 965
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGC 600

Qy    966 tatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactgg 1025
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGG 660

Qy   1026 aacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacga 1085
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 AACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACGA 720

Qy   1086 tctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgtttta 1145
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTA 780

Qy   1146 tgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatgga 1205
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGA 840

Qy   1206 caatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatgac 1265
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGAC 900

Qy   1266 ttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggagaa 1325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAA 960

Qy   1326 catcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccgt 1385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGT 1020

Qy   1386 gttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagccca 1445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCCA 1080

Qy   1446 cgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcgt 1505
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGT 1140

Qy   1506 gctgacagtggccttcatcaccttaaatgttgtgcttatggctaccggaaatgcttggg 1565
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGG 1200

Qy   1566 gaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtgggaa 1625
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGAA 1260

Qy   1626 ataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgt 1685
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 ATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTGT 1320

Qy   1686 taaattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtgc 1745
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTGC 1380

Qy   1746 tttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaata 1805
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAATA 1440

Qy   1806 tgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggacg 1865
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1441 TGTATCGTGCCCCCTCTGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGACG 1500

Qy   1866 tgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttgggag 1925
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGAG 1560

Qy   1926 tgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatcttat 1985
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   1561 TGCGGGATTCAAAGGTGGTCCCACGGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTAT 1620

Qy   1986 cttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatct 2045
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCTCATCT 1680

Qy   2046 ctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacctttag 2105
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CTGTCGTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTTAG 1740

Qy   2106 agtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttcttgc 2165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTTCTTGC 1800

Qy   2166 tttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagatgatcat 2225
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATAATCAT 1860

Qy   2226 tgcttatgccaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgctgg 2285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGTGCTGG 1920

Qy   2286 gcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||
Db   1921 GCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 1955
```

```
RESULT  10
HUMHUGBR2
LOCUS       HUMHUGBR2               2368 bp    mRNA    linear   PRI 07-MAR-1995
DEFINITION  Human bilirubin UDP-glucuronosyltransferase isozyme 2 mRNA,
            complete cds.
ACCESSION   M57951
VERSION     M57951.1  GI:184474
KEYWORDS    UDP-glucuronosyltransferase; bilirubin UDP-glucuronosyltransferase.
SOURCE      Human adult female liver, cDNA to mRNA, clones Z17 and Z17P5E2.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2368)
  AUTHORS   Ritter,J.K., Crawford,J.M. and Owens,I.S.
  TITLE     Cloning of two human liver bilirubin UDP-glucuronosyltransferase
            cDNAs with expression in COS-1 cells
  JOURNAL   J. Biol. Chem. 266 (2), 1043-1047 (1991)
  MEDLINE   91093210
FEATURES             Location/Qualifiers
     source          1. .2368
                     /organism="Homo sapiens"
                     /isolate="AK"
                     /db_xref="taxon:9606"
                     /map="Unassigned"
                     /clone="Z17"
                     /sex="female"
                     /tissue_type="liver"
                     /dev_stage="adult"
                     /tissue_lib="lambda-ZAP"
     gene            1. .2368
                     /gene="UGT2B"
     CDS             30. .1634
                     /gene="UGT2B"
                     /standard_name="bilirubin UDP-glucuronosyltransferase
                     isozyme 2"
                     /EC_number="2.4.1.17"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 2 family polypeptide
                     B"
                     /protein_id="AAA63196.1"
                     /db_xref="GI:184475"
                     /db_xref="GDB:G00-127-753"
                     /translation="MARGLQVPLPRLATGLLLLLSVQPWAESGKVLVVPTDGSPWLSM
                     REALRELHARGHQAVVLTPEVNMHIKEEKFFTLTAYAVPWTQKEFDRVTLGYTQGFFE
                     TEHLLKRYSRSMAIMNNVSLALHRCCVELLHNEALIRHLNATSFDVVLTDPVNLCGAV
```

```
Qy   1502 tcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgct 1561
          |||||||||||||||||||||| ||||||||||||||| ||||||||||||||||||||
Db   1543 TCGTGCTGACAGTGGCCTTCATCGCCTTTAAATGTTGTGCCTATGGCTACCGGAAATGCT 1602

Qy   1562 tggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtg 1621
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   1603 TCGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTTGGTG 1662

Qy   1622 ggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatca 1681
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   1663 GGAAATAAGGTAAAATTTTGAACCATTCCCTAGTCAGTTCCAAACTTGAAAACAGAATCA 1722

Qy   1682 gtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccaga 1741
          |||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db   1723 GTGTTAAATTCATTTTATTC---TTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGA 1779

Qy   1742 gtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttg 1801
          |||||||||||||||| |||||||||||||||||| || ||||||||||||||||||||
Db   1780 GTGCTTTAAAAAATTATCTTAAATAAAAATAATAGAATCACTAGTCAGTAAAGATATTTG 1839

Qy   1802 aatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagag 1861
          |||||||||| || ||||||||||||||| ||||| ||||||| |||||| ||||||||
Db   1840 AATATGTATCATG-CCCCTCCGGTGTCTTCGATCAAGATGACACATGCCATCTTTCAGAG 1898

Qy   1862 gacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttg 1921
          || |||||||||||||||| ||||| ||  |||||||||||| |||||||| ||||||||
Db   1899 GATGTGCAGACAGGCTGGCGTTCTACATGCCTTTTCTTACTCGGAAACATGACCTGTTTG 1958

Qy   1922 ggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatc 1981
          ||||||| |||||||||||||||||||||||||||| ||||| |||||||||||||||||||
Db   1959 GGAGTGCAGGATTCAAAGGTGGTCCCACCGCTGTCCCTATTGCAAATGGCAGTTTTAATC 2018

Qy   1982 ttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctc 2041
          |||||||||||||||| |||||||||| |||||||||||||||||||||||  || |||
Db   2019 TTATCTTTTGGCTTCTGTAGATGGTTGCGATTGATCCTTAACCAATAATGGTTGGTTCTC 2078

Qy   2042 atctctgtcc-tgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacc 2100
          |||||||||| |||||||||| ||| ||||||||||||||||||||| |||||||||||
Db   2079 ATCTCTGTCCGTGCTTCATAGGCGCCCCCTTGTGTGTTTAAAGAAGGGAAACTTTGTACC 2138

Qy   2101 tttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgattt 2160
          |||||||  |||||||||||| ||||||||||||||||||||||||||||||| ||| |
Db   2139 TTTAGAGCATAGGTGAAATGAGTGAATGGCTTGGAGTGCACTGAGAACAGCATACGATCT 2198

Qy   2161 cttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagata 2220
          ||||||||| |||||||||||||||||||| | |||||| |||||||||| ||||||||||
Db   2199 ATTGCTTTGAGGAAAAAGAATGATGCTGTAAAATTGATGGGTGGTGTGTTTGAGAAGATA 2258

Qy   2221 atcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagt 2280
          |||||||||||||||||||||||||||||||| ||||||||||||||||||  |||||||
Db   2259 ATCATTGCTTATGTCAAATGGAGCTGAATTTCATAAAAACCCAAAATACAGTGATGAAGT 2318

Qy   2281 gctgggcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||| || ||||||||| ||||||||| |||||
Db   2319 GCTGGGCAAGCGTA-TTTTTTTCTTATGTTTCCTGCAACT 2357


RESULT   9
AK025403
LOCUS       AK025403               2003 bp    mRNA    linear   PRI 29-SEP-2000
DEFINITION  Homo sapiens cDNA: FLJ21750 fis, clone COLF5999, highly similar to
            AF056188 Homo sapiens UDP-glucuronosyltransferase 1A9 (UGT1A9)
            mRNA.
ACCESSION   AK025403
VERSION     AK025403.1  GI:10437911
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens colon mucosa cDNA to mRNA, clone_lib:ColF
            clone:COLF5999.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
            Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
            Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 2003)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
            Sugano, Institute of Medical Science, University of Tokyo,
            Laboratory of Genome Structure Analysis, Human Genome Center;
            Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Departent of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1. .2003
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="COLF5999"
                     /clone_lib="ColF"
                     /tissue_type="colon mucosa"
                     /note="cloning vector pME18SFL3"
     misc_feature    1. .2003
                     /note="highly similar to AF056188 Homo sapiens
                     UDP-glucuronosyltransferase 1A9 (UGT1A9) mRNA"
BASE COUNT      551 a    414 c    445 g    593 t
ORIGIN


  Query Match           81.7%;  Score 1895.8;  DB 9;  Length 2003;
  Best Local Similarity   98.1%;  Pred. No. 0;
  Matches 1918;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy    366 ttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaata 425
          | ||||||||||||||||  |||||||||||||||||||| |||  ||||||||||||||||
Db      1 TCTTGACTTATTTTTTTCGCATTGCAGGAGTTTGTTTAATGACCGAAAATTAGTAGAATA 60

Qy    426 cttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggcttaat 485
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||| ||||||||||
Db     61 CTTAAAGGAGAGTTCTTTTGATGCAGTGTTTCTGGATCCTTTTGATACCTGTGGCTTAAT 120

Qy    486 tgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgccacta 545
          |||||| |||||||||||||||||| ||||||||| |||||||||| |||||||| |
Db    121 TGTTGCTAAATATTTCTCCCTCCCCTCTGTGGTCTTCACCAGGGGAATATTTTGCCACCA 180

Qy    546 tcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttagg 605
          ||||||||||||||||||||||||||||||||||||||||||||||||||    |||||||||
Db    181 TCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAATGATCTCTTAGG 240

Qy    606 gttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggagga 665
          |||||||||||||||||||||||||||||||||||| |||||||||||  ||||||||||||
Db    241 GTTCTCAGATGCCATGACTTTCAAGGAGAGAGTATGGAACCACATCGTGCACTTGGAGGA 300

Qy    666 acatttattatgccaccgtttttttcaaaaatgccctagaaatagcctctgaaattctcca 725
          |||||||| |||||   | |||| | ||||||||||||||||||||||||||||||||||||
Db    301 CCATTTATTTTGCCAGTATCTTTTTAGAAATGCCCTAGAAATAGCCTCTGAAATTCTCCA 360

Qy    726 aacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacgga 785
          ||| ||||| |||| ||||||||||||| ||||||||||||||||||||||||||||||||
Db    361 AACCCCTGTCACGGCATATGATCTCTACAGTCACACATCAATTTGGTTGTTGCGAACGGA 420

Qy    786 ctttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaa 845
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTTTGTTTTGGACTATCCCAAACCCGTGATGCCCAACATGATCTTCATTGGTGGTATCAA 480
```

```
VERSION     AF104336.1  GI:6537137
KEYWORDS    .
SOURCE      crab-eating macaque.
  ORGANISM  Macaca fascicularis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 2396)
  AUTHORS   Albert,C., Vallee,M., Beaudry,G., Belanger,A. and Hum,D.W.
  TITLE     The monkey and human uridine diphosphate-glucuronosyltransferase
            UGT1A9, expressed in steroid target tissues, are
            estrogen-conjugating enzymes
  JOURNAL   Endocrinology 140 (7), 3292-3302 (1999)
  MEDLINE   99312152
   PUBMED   10385426
REFERENCE   2  (bases 1 to 2396)
  AUTHORS   Albert,C., Vallee,M., Beaudry,G., Belanger,A. and Hum,D.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-1998) Molecular Endocrinology, Laval University,
            2705 Laurier Boulevard, Sainte-Foy, Que G1V4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .2396
                     /organism="Macaca fascicularis"
                     /db_xref="taxon:9541"
     CDS             60. .1652
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase UGT1A09"
                     /protein_id="AAF15546.1"
                     /db_xref="GI:6537138"
                     /translation="MTRTGWSSPLPLCVCLLLACGFAEAGKLLVVPVDGSHWLSMRSV
                     VEKLILRGHEVVVVMPKVSWQLGRPLNCTVKTYSTSYTLEDLDREFKAFRQAQWKAQV
                     QSIFSLIMSSYNDIFDLFFSNCRSLFKDKKLVEYLKESSFDAVFLDPFDTCGLIVAKY
                     LSLPSVVFARGVFCHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEERL
                     LCHRFFKSALEIASEILQTPVTVYDLHRQISIWLLRMDFVLDYPKPVMPNMVFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVAEIPEKKAMAIADALGKIPQTVLWRY
                     TGTPPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGIYEGICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMHLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAIVLTVAFIAFKCCA
                     YGYRKCFGKKGRVKKAHKSKTH"
BASE COUNT      645 a    513 c    558 g    680 t
ORIGIN
```

```
  Query Match           89.5%;  Score 2076;  DB 9;  Length 2396;
  Best Local Similarity  94.6%;  Pred. No. 0;
  Matches 2194;  Conservative    0;  Mismatches  120;  Indels    6;  Gaps    4;

Qy       2 tcagctgcagttctctgatggcttgcacagggtggaccagccccctttcctctatgtgtgt 61
           ||  |||| |||| |||||| || |||||||||||| ||| |||||||| ||||||||||
Db      43 TCGGCTGCGGTTCCCTGATGACTCGCACAGGGTGGAGCAGTCCCCTTCCCCTATGTGTGT 102

Qy      62 gtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatg 121
           ||||  |||||| |||||||||||||||||||||||||||||| |||||||||||| ||||||
Db     103 GTCTGCTGCTGGCCTGTGGCTTTGCCGAGGCAGGGAAGCTGCTGGTAGTGCCCGTGGATG 162

Qy     122 ggagccactggttcaccatgaggtcggtggtggagaaactcattctcagggggcatgagg 181
           ||||||||||| || ||||||||||||||||||||||||||||||||||||||||||||||
Db     163 GGAGCCACTGGCTCTCCATGAGGTCGGTGGTGGAGAAACTCATTCTCAGGGGGCATGAGG 222

Qy     182 tggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtga 241
           |||| |||||||||||| |||||||||||||| ||||||||| ||||||||||||||||||
Db     223 TGGTCGTAGTCATGCCAAAGGTGAGTTGGCAGCTGGGAAGACCACTGAATTGCACAGTGA 282

Qy     242 agacttattcaacttcatataccctggaggatctggaccgggagttcaaggctttgccc 301
           |||||||||||||||||| || ||||||||||||||||||||||||||||||||||  ||
Db     283 AGACTTATTCAACTTCATACACTCTGGAGGATCTGGACCGGGAGTTCAAGGCTTTTCGCC 342

Qy     302 atgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatg 361
           | |||||||||||||||||||||| |||||||| ||||||| ||||| |||| |||||||
Db     343 AGGCTCAATGGAAAGCACAAGTACAAAGTATATTTTCTCTAATAATGAGTTCCTACAATG 402

Qy     362 acatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtag 421
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||| |||||||
```

```
Db     403 ACATTTTTGACTTATTTTTTTCAAATTGCAGGAGTTTGTTTAAGGACAAAAAGTTAGTAG 462

Qy     422 aatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggct 481
           |||||||||||||||||||||||||||||| || ||||| |||||||||||| ||||||||
Db     463 AATACTTAAAGGAGAGTTCTTTTGATGCGGTATTTCTGGATCCTTTTGATACCTGTGGCT 522

Qy     482 taattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgcc 541
           |||| |||||||||||| |||||||||||||| |||||||||||||||||||| || ||||||
Db     523 TAATCGTTGCCAAATATCTCTCCCTCCCCTCTGTGGTCTTCGCCAGGGGAGTATTTTGCC 582

Qy     542 actatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctct 601
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     583 ACTATCTTCAAGAAGGTGCGCAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTCT 642

Qy     602 tagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttgg 661
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     643 TAGGGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTACGGAACCACATCATGCACTTGG 702

Qy     662 aggaacatttattatgccaccgtttttttcaaaaatgccctagaaatagcctctgaaattc 721
           |||||| |||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     703 AGGAACGTTTATTATGCCACCGTTTTTTCAAAAGTGCCCTAGAAATAGCCTCTGAAATTC 762

Qy     722 tccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaa 781
           | |||||||||| || |  |||||||  |||| || | |||||||||||||| ||||||
Db     763 TTCAAACACCTGTCACAGTATATGATCTTCACAGACAAATATCAATTTGGTTGCTGCGAA 822

Qy     782 cggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggta 841
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     823 TGGACTTTGTTTTGGACTATCCCAAACCCGTGATGCCCAACATGGTCTTCATTGGTGGTA 882

Qy     842 tcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctg 901
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     883 TCAACTGCCATCAGGGAAAGCCATTGCCTATGGAATTTGAAGCCTACATTAATGCTTCTG 942

Qy     902 gagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaaga 961
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     943 GAGAACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCGCAGAAATTCCAGAGAAGA 1002

Qy     962 aagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtaca 1021
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1003 AAGCTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACA 1062

Qy    1022 ctggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaa 1081
           |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1063 CTGGAACCCCACCATCGAATCTTGCGAACAATACGATACTTGTTAAGTGGCTACCCCAAA 1122

Qy    1082 acgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtg 1141
           ||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1123 ACGATCTTCTTGGTCACCCAATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTA 1182

Qy    1142 tttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcaga 1201
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1183 TTTATGAAGGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGA 1242

Qy    1202 tggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaa 1261
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    1243 TGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTTGAAA 1302

Qy    1262 tgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaagg 1321
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    1303 TGACTTCTGAAGATTTAGAAAATGCCCTAAAAGCAGTCATCAATGACAAAAGCTACAAGG 1362

Qy    1322 agaacatcagcgcctctccagccttcacaaggaccgcccggtggagccgctggacctgg 1381
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1363 AGAACATCATGCACCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGG 1422

Qy    1382 ccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcag 1441
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    1423 CCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCACCACACCTGCGCCCCGCAG 1482

Qy    1442 cccacgaccctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccg 1501
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||| |||||
Db    1483 CCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATCGGTTTCCTCCTGGCCA 1542
```

```
Db    265 ATTCAACTTCATATA-CCTGGAGGATCTGGACCGGGAGTTCAAGGCTTTTGCCCATGCTC 323

Qy    308 aatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgacattt 367

Db    324 AATGGAAAGCACAAGTACGAAGTATATATTCTCTATTAATGGGTTCATACAATGACATTT 383

Qy    368 ttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatact 427

Db    384 TTGACTTATTTTTTTCAAATTGCAGGAGTTTGTTTAAAGACAAAAAATTAGTAGAATACT 443

Qy    428 taaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggcttaattg 487

Db    444 TAAAGGAGAGTTCTTTTGATGCAGTGTTTCTCGATCCTTTTGATAACTGTGGCTTAATTG 503

Qy    488 ttgccaaatatttctccctccccctccgtggtcttcgccaggggaatactttgccactatc 547

Db    504 TTGCCAAATATTTCTCCCTCCCCTCCGTGGTCTTCGCCAGGGGAATACTTTGCCACTATC 563

Qy    548 ttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttagggt 607

Db    564 TTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTCTTAGGGT 623

Qy    608 tctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggaggaac 667

Db    624 TCTCAGATGACATGACTTTCAAGGAGAGAGTACGGAACCACATCATGCACTTGGAGGAAC 683

Qy    668 atttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaattctccaaa 727

Db    684 ATTTATTATGCCACCGTTTTTTCAAAAATGCCCTAGAAATAGCCTCTGAAATTCTCCAAA 743

Qy    728 cacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacggact 787

Db    744 CACCTGTTACGGAGTATGATCTCTACAGCCACACATCAATTTGGTTGTTGCGAACGGACT 803

Qy    788 ttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaact 847

Db    804 TTGTTTTGGACTATCCCAAACCCGTGATGCCCAACATGATCTTCATTGGTGGTATCAACT 863

Qy    848 gccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctggagaac 907

Db    864 GCCATGAGAGGAAAGCGTTGCCTATGGAATTTGAAGCCTACATTAATGCTTCTGGAGAAC 923

Qy    908 atggaattgtggttttctctttgggatcaatggtctcagaaattccagagaagaaagcta 967

Db    924 ATGGAATTGTGGGTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTA 983

Qy    968 tggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactggaa 1027

Db    984 TGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGGAA 1043

Qy   1028 cccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacgatc 1087

Db   1044 CCCGACCATGCAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACGATC 1103

Qy   1088 tgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgtttatg 1147

Db   1104 TGCTTGGTCACCCGATGACCAGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATG 1163

Qy   1148 aaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatggaca 1207

Db   1164 AAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACA 1223

Qy   1208 atgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatgactt 1267

Db   1224 ATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCATGAATGTTCTGGAAATGACTT 1283

Qy   1268 ctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggagaaca 1327

Db   1284 CTGAAGAATTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAACA 1343

Qy   1328 tcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccgtgt 1387

Db   1344 TCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGT 1403

Qy   1388 tctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagcccacg 1447

Db   1404 TCTGGGTGGAGTTTGTTATGAGGCACAAGGGCGCGACACACCTGCGCCCCGCAGCCCACG 1463

Qy   1448 acctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcgtgc 1507

Db   1464 ACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGCCGTGC 1523

Qy   1508 tgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttgggga 1567

Db   1524 TGACAGTGGCCTTCATCACCTGTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGGA 1583

Qy   1568 aaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtgggaaat 1627

Db   1584 AAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAGGTGGGTGGGAAAT 1643

Qy   1628 aaggtaaaatttttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgtta 1687

Db   1644 AAGGTAAAATTTTGAACCA-TCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTGTTA 1702

Qy   1688 aattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtg-ct 1746

Db   1703 AATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTGTCT 1762

Qy   1747 ttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaatat 1806

Db   1763 TTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAATAT 1822

Qy   1807 gtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggacgt 1866

Db   1823 GTATCGTGCCCCCTCCGGAGTCTTTGATCAGGATGACATGTGCCATTTTTCAAAGGACGT 1882

Qy   1867 gcagacaggctggcattc-tagattacttttcttactctgaaacatggcctgtttgggag 1925

Db   1883 GCAGACAGGCTGGCATTCTTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGAG 1942

Qy   1926 tgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatcttat 1985

Db   1943 TGCGGGATTCAAAGGTGGTCCCACGGCTGCCCCTACTGCAAATGGGAGTTTTAATCCTAT 2002

Qy   1986 cttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatct 2045

Db   2003 CTTTTGGCTTCTGCAGATGGTTGC-ATTGATCCTTAACCAATAATGGTCAGTCCTCATCT 2061

Qy   2046 ctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacctttag 2105

Db   2062 CTGTCCTGACTTCATAGGTGCCAGGTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTTAG 2121

Qy   2106 agtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttcttgc 2165

Db   2122 AGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTTCTTGC 2181

Qy   2166 tttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatcat 2225

Db   2182 TTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTTTATTTGAGAAGATAATCAT 2241

Qy   2226 tgcttatgtcaaatggagctgaatttgataaaaaacccaaaatacagctatgaagtgctgg 2285

Db   2242 TGCTTATGTCAAATGGAGGTTAATTTGATAAAAACCCAAAATACAGCTATGAAGTGCTGG 2301

Qy   2286 gcaagtttac-tttttttctgatgtttcctacaact 2320

Db   2302 GCAAGTTTACTTTTTTTTCTGATGTTTCCTACAAAT 2337
```

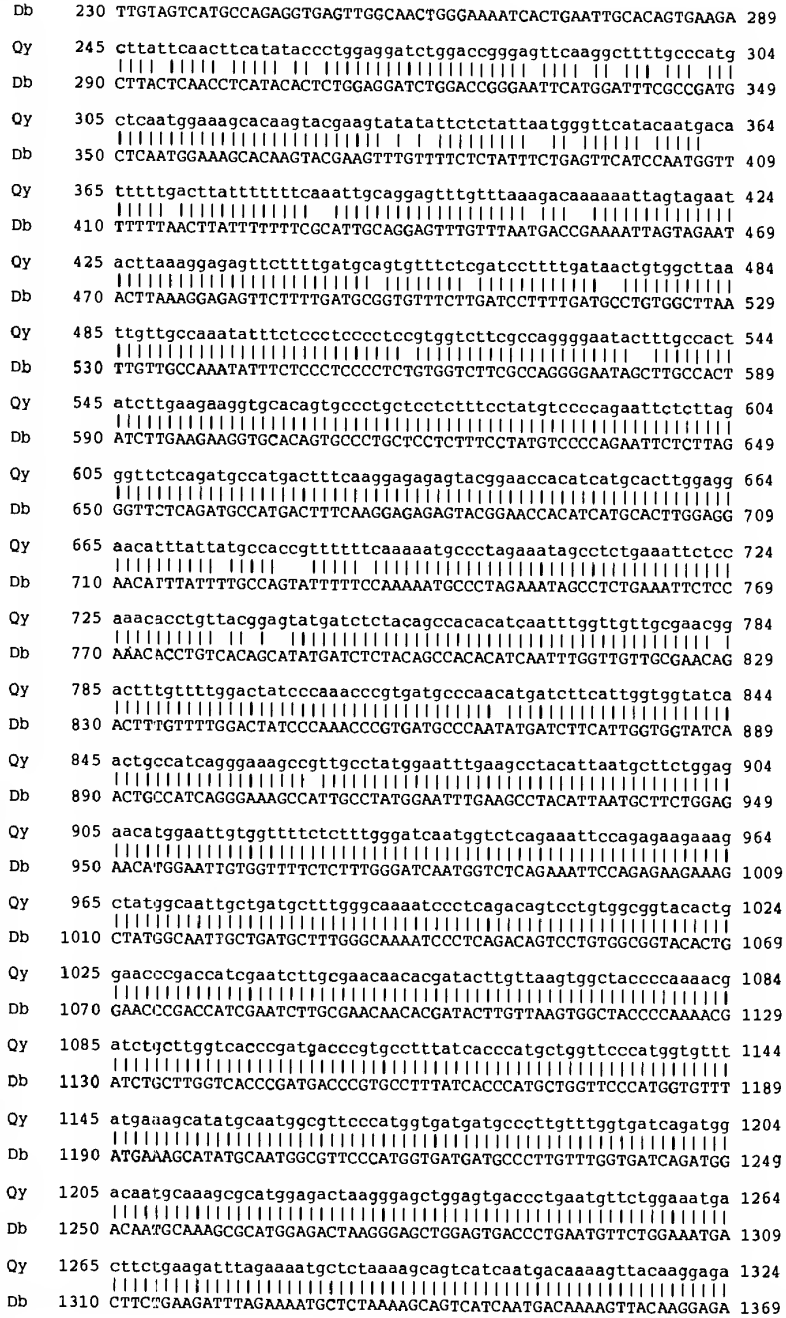
RESULT 8
AF104336
LOCUS AF104336 2396 bp mRNA linear PRI 13-DEC-1999
DEFINITION Macaca fascicularis UDP-glucuronosyltransferase UGT1A09 mRNA, complete cds.
ACCESSION AF104336

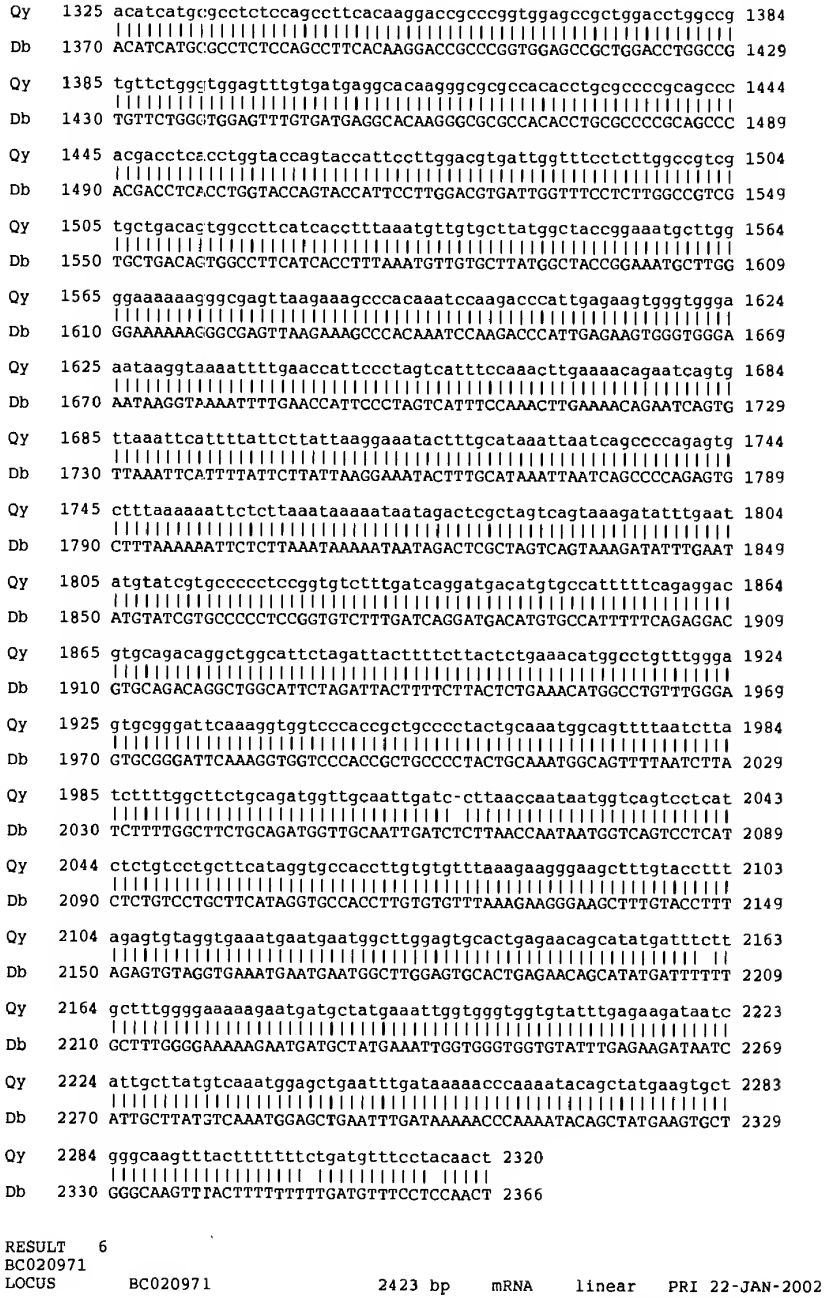
```
Db   1196 ATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGG 1255

Qy   1205 acaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatga 1264
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1256 ACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGA 1315

Qy   1265 cttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggaga 1324
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1316 CTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGA 1375

Qy   1325 acatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccg 1384
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1376 ACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCG 1435

Qy   1385 tgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagccc 1444
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1436 TGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCC 1495

Qy   1445 acgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcg 1504
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1496 ACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCG 1555

Qy   1505 tgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttgg 1564
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1556 TGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGG 1615

Qy   1565 ggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtggga 1624
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1616 GGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGA 1675

Qy   1625 aataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtg 1684
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1676 AATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTG 1735

Qy   1685 ttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtg 1744
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1736 TTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTG 1795

Qy   1745 cttt-aaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaa 1803
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1796 CTTTAAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAA 1855

Qy   1804 tatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagagga 1863
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1856 TATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGA 1915

Qy   1864 cgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttggg 1923
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1916 CGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGG 1975

Qy   1924 agtgcgggattcaaaggtggtcccaccgctgccctactgcaaatggcagttttaatctt 1983
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1976 AGTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAATCTT 2035

Qy   1984 atcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcat 2043
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2036 ATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCTCAT 2095

Qy   2044 ctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacctttt 2103
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2096 CTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTT 2155

Qy   2104 agagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttctt 2163
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2156 AGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTTCTT 2215

Qy   2164 gctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatc 2223
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2216 GCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATAATC 2275

Qy   2224 attgcttatgtcaaatggagctgaatttgataaaaaacccaaaatacagctatgaagtgct 2283
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2276 ATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGTGCT 2335
```

```
Qy   2284 gggcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||||
Db   2336 GGGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 2372


RESULT   7
S55985
LOCUS       S55985                  2392 bp    mRNA    linear   PRI 04-MAY-2000
DEFINITION  Homo sapiens UDP-glucuronosyltransferase mRNA, complete cds.
ACCESSION   S55985
VERSION     S55985.1  GI:235453
KEYWORDS    .
SOURCE      human liver.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2392)
  AUTHORS   Wooster,R., Sutherland,L., Ebner,T., Clarke,D., Da Cruz e Silva,O.
            and Burchell,B.
  TITLE     Cloning and stable expression of a new member of the human liver
            phenol/bilirubin: UDP-glucuronosyltransferase cDNA family
  JOURNAL   Biochem. J. 278 (Pt 2), 465-469 (1991)
  MEDLINE   91378895
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 55985] from the original journal article.
FEATURES             Location/Qualifiers
     source          1. .2392
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
     CDS             35. .1627
                     /note="phenol/bilirubin; lugP4; conceptual translation
                     presented here differs from translation in publication"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase"
                     /protein_id="AAB19791.2"
                     /db_xref="GI:7690346"
                     /translation="MACTGWTSPLPLCVCLLLTCGFAEAGKLVVVPMDGSHWFTMRSV
                     VEKLILRGHEVVVVSARGELATGKINELHSEDLFNFIYLEDLDREFKAFAHAQWKAQV
                     RSIYSLLMGSYNDIFDLFFSNCRSLFKDKKLVEYLKESSFDAVFLDPFDNCGLIVAKY
                     FSLPSVVFARGILCHYLEEGAQCPAPLSYVPRILLGFSDDMTFKERVRNHIMHLEEHL
                     LCHRFFKNALEIASEILQTPVTEYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHERKALPMEFEAYINASGEHGIVGFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPCNLANNTILVKWLPQNDLLGHPMTSAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTMNVLEMTSEELENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGATHLRPAAHDLTWYQYHSLDVIGFLLAAVLTVAFITCKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT      661 a    492 c    554 g    685 t
ORIGIN
```

```
  Query Match             93.0%;  Score 2156.8;  DB 9;  Length 2392;
  Best Local Similarity   97.7%;  Pred. No. 0;
  Matches 2262;  Conservative    0;  Mismatches   47;  Indels    7;  Gaps    7;

Qy       9 cagttctctgatggcttgcacagggtggaccagccccctcctctatgtgtgtgtctgct 68
           ||  |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25 CACCTCTGGCATGGCTTGCACAGGGTGGACCAGCCCCCTTCCTCTATGTGTGTCTGCT 84

Qy      69 gctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatgggagcca 128
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      85 GCTGACCTGTGGCTTTGCCGAGGCAGGGAAGCTAGTGGTAGTGCCCATGGATGGGAGCCA 144

Qy     129 ctggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgaggtggttgt 188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145 CTGGTTCACCATGAGGTCGGTGGTGGAGAAACTCATTCTCAGGGGGCATGAGGTGGTTGT 204

Qy     189 agtca-tgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaagactt 247
           ||||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     205 AGTCAGTGCCAGAGGTGAGTTGGCAACTGGGAAGATCAATGAATTGCACAGTGAAGACTT 264

Qy     248 attcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcccatgctc 307
```

```
DEFINITION  Homo sapiens, Similar to UDP glycosyltransferase 1 family,
            polypeptide A9, clone MGC:9353 IMAGE:3850654, mRNA, complete cds.
ACCESSION   BC020971
VERSION     BC020971.1  GI:18088086
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2423)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 20 Row: m Column: 8
            This clone was selected for full length sequencing because it
            passed the following selection criteria: GenomeScan gene
            prediction, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1. .2423
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:9353 IMAGE:3850654"
                     /tissue_type="Colon, adenocarcinoma"
                     /clone_lib="NIH_MGC_65"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             70. .1662
                     /codon_start=1
                     /product="Similar to UDP glycosyltransferase 1 family,
                     polypeptide A9"
                     /protein_id="AAH20971.1"
                     /db_xref="GI:18088087"
                     /translation="MARAGWTSPVPLCVCLLLTCGFAEAGKLLVVPMDGSHWFTMQSV
                     VEKLILRGHEVVVVMPEVSWQLERSLNCTVKTYSTSYTLEDQNREFMVFAHAQWKAQA
                     QSIFSLLMSSSSGFLDLFFSHCRSLFNDRKLVEYLKESSFDAVFLDPFDTCGLIVAKY
                     FSLPSVVFTRGIFCHHLEEGAQCPAPLSYVPNDLLGFSDAMTFKERVWNHIVHLEDHL
                     FCQYLFRNALEIASEILQTPVTAYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT      643 a    514 c    559 g    707 t
ORIGIN


  Query Match             94.6%;  Score 2194.6;  DB 9;  Length 2423;
  Best Local Similarity   97.0%;  Pred. No. 0;
  Matches 2247;  Conservative    0;  Mismatches   69;  Indels    1;  Gaps    1;

Qy       5 gctgcagttctctgatggcttgcacagggtggaccagccccctcctctatgtgtgtgtc 64
           ||||||||||||| |||||| || ||||||||||||||||| ||||| ||||||||||||
Db      56 GCTGCAGTTCTCTCATGGCTCGCGCAGGGTGGACCAGCCCCGTTCCTTTATGTGTGTGTC 115

Qy      65 tgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatggga 124
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     116 TACTGCTGACCTGTGGCTTTGCCGAGGCAGGGAAGCTGCTGGTAGTGCCCATGGATGGGA 175

Qy     125 gccactggttcaccatgaggtcggtggtggagaaactcattctcagggggcatgaggtgg 184
           | |||||||||||||||  |||||||||||||||||| || |||||||||||||||||||
Db     176 GTCACTGGTTCACCATGCAGTCGGTGGTGGAGAAACTTATCCTCAGGGGGCATGAGGTGG 235

Qy     185 ttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaaga 244
           ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     236 TTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGAAAGATCACTGAATTGCACAGTGAAGA 295

Qy     245 cttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcccatg 304
           |||| ||||| || || || ||||| |||| | ||||||| |||| || ||| ||||||||
Db     296 CTTACTCAACCTCGTACACTCTGGAAGATCAGAACCGGGAATTCATGGTTTTCGCCCATG 355

Qy     305 ctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgaca 364
           |||||||||||||||| | || |||||||| |||||||||||| |||||| || ||
Db     356 CTCAATGGAAAGCACAGGCACAAAGTATATTTTCTCTATTAATGAGTTCATCCAGTGGTT 415

Qy     365 tttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaat 424
           || ||||||||||||||  |||||||||||||||||| |||  ||||||||||||||
Db     416 TTCTTGACTTATTTTTTTCGCATTGCAGGAGTTTGTTTAATGACCGAAAATTAGTAGAAT 475

Qy     425 acttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggcttaa 484
           |||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||
Db     476 ACTTAAAGGAGAGTTCTTTTGATGCAGTGTTTCTGGATCCTTTTGATACCTGTGGCTTAA 535

Qy     485 ttgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgccact 544
           ||||||| |||||||||||||||||||| |||||||| |||||||||| ||||||||
Db     536 TTGTTGCTAAATATTTCTCCCTCCCCTCTGTGGTCTTCACCAGGGGAATATTTTGCCACC 595

Qy     545 atcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttag 604
           |||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||
Db     596 ATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAATGATCTCTTAG 655

Qy     605 ggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggagg 664
           |||||||||||||||||||||||||||||||||||| |||||||||| ||||||||||
Db     656 GGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTATGGAACCACATCGTGCACTTGGAGG 715

Qy     665 aacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaattctcc 724
           | |||||||| |||||   | |||| | |||||||||||||||||||||||||||||||
Db     716 ACCATTTATTTTGCCAGTATCTTTTTAGAAATGCCCTAGAAATAGCCTCTGAAATTCTCC 775

Qy     725 aaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacgg 784
           |||| |||||  |||  |||||||||||||| ||||||||||||||||||||||||||||
Db     776 AAACCCCTGTCACGGCATATGATCTCTACAGTCACACATCAATTTGGTTGTTGCGAACGG 835

Qy     785 actttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatca 844
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     836 ACTTTGTTTTGGACTATCCCAAACCCGTGATGCCCAACATGATCTTCATTGGTGGTATCA 895

Qy     845 actgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctggag 904
           |||| ||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     896 ACTGTCATCAGGGAAAGCCATTGCCTATGGAATTTGAAGCCTACATTAATGCTTCTGGAG 955

Qy     905 aacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaagaaaag 964
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     956 AACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAG 1015

Qy     965 ctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactg 1024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1016 CTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTG 1075

Qy    1025 gaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacg 1084
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1076 GAACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACG 1135

Qy    1085 atctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgttt 1144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1136 ATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTT 1195

Qy    1145 atgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatgg 1204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```





RESULT 6
BC020971
LOCUS BC020971 2423 bp mRNA linear PRI 22-JAN-2002

```
Qy   1205 acaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatga 1264
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1220 ACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGA 1279

Qy   1265 cttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggaga 1324
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1280 CTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGA 1339

Qy   1325 acatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccg 1384
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1340 ACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCG 1399

Qy   1385 tgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagccc 1444
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1400 TGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCC 1459

Qy   1445 acgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcg 1504
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1460 ACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCG 1519

Qy   1505 tgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttgg 1564
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1520 TGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGG 1579

Qy   1565 ggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtggga 1624
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1580 GGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGA 1639

Qy   1625 aataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtg 1684
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1640 AATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTG 1699

Qy   1685 ttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtg 1744
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1700 TTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTG 1759

Qy   1745 ctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaat 1804
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1760 CTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAAT 1819

Qy   1805 atgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggac 1864
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1820 ATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGAC 1879

Qy   1865 gtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttggga 1924
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1880 GTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGA 1939

Qy   1925 gtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatctta 1984
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1940 GTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTA 1999

Qy   1985 tcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatc 2044
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2000 TCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCTCATC 2059

Qy   2045 tctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtaccttta 2104
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2060 TCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTTA 2119

Qy   2105 gagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttcttg 2164
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   2120 GAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTTTTTG 2179

Qy   2165 ctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagatgtaatca 2224
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2180 CTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATAATCA 2239

Qy   2225 ttgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgctg 2284
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2240 TTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGTGCTG 2299

Qy   2285 ggcaagtttacttttttttctgatgtttcctacaact 2320
          ||||||||||||||||||| |||||||||||| |||||
Db   2300 GGCAAGTTTACTTTTTTTTTGATGTTTCCTCCAACT 2335
```

```
RESULT   5
AF462268
LOCUS       AF462268               2394 bp    mRNA    linear   PRI 07-FEB-2002
DEFINITION  Homo sapiens UDP-glucuronosyltransferase UGT1A8*1 (UGT1A8) mRNA,
            UGT1A8*1 allele, complete cds.
ACCESSION   AF462268
VERSION     AF462268.1  GI:18568218
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2394)
  AUTHORS   Huang,Y.-H., Nguyen,N., Geske,D. and Tukey,R.
  TITLE     Identification and functional characterization of UGT1A8*1,
            UGT1A8*2 and UGT1A8*3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2394)
  AUTHORS   Huang,Y.-H., Nguyen,N., Geske,D. and Tukey,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2001) Pharmacology, UCSD, 9500 Gilman Drive, La
            Jolla, CA 92093, USA
FEATURES             Location/Qualifiers
     source          1. .2394
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="colon"
     gene            1. .2394
                     /gene="UGT1A8"
                     /allele="UGT1A8*1"
     CDS             64. .1656
                     /gene="UGT1A8"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase UGT1A8*1"
                     /protein_id="AAL75964.1"
                     /db_xref="GI:18568219"
                     /translation="MARTGWTSPIPLCVSLLLTCGFAEAGKLLVVPMDGSHWFTMQSV
                     VEKLILRGHEVVVVMPEVSWQLGKSLNCTVKTYSTSYTLEDLDREFMDFADAQWKAQV
                     RSLFSLFLSSSNGFFNLFFSHCRSLFNDRKLVEYLKESSFDAVFLDPFDACGLIVAKY
                     FSLPSVVFARGIACHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHL
                     FCQVFSKNALEIASEILQTPVTAYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT      624 a    513 c    555 g    702 t
ORIGIN
```

```
  Query Match             95.1%;  Score 2207.4;  DB 9;  Length 2394;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 2255;  Conservative    0;  Mismatches   61;  Indels    1;  Gaps    1;

Qy       5 gctgcagttctctgatggcttgcacagggtggaccagccccctcctctatgtgtgtgtc 64
           |||||||||||| |||||| |||||||||||||||||| |||| |||||||| | ||
Db      50 GCTGCAGTTCTCTCATGGCTCGCACAGGGTGGACCAGCCCCATTCCCCTATGTGTTCTC 109

Qy      65 tgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatggga 124
           |||||||||||||||||||| |||||||||||||| ||||||||||||||||||||||||
Db     110 TGCTGCTGACCTGTGGCTTTGCTGAGGCAGGGAAGCTGCTGGTAGTGCCCATGGATGGGA 169

Qy     125 gccactggttcaccatgaggtcggtggtggagaaactcattctcaggggggcatgaggtgg 184
           | ||||||||||||||  |||||||||||||||||| || ||||||||||||||||||||
Db     170 GTCACTGGTTCACCATGCAGTCGGTGGTGGAGAAACTTATCCTCAGGGGGCATGAGGTGG 229

Qy     185 ttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaaga 244
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
```

```
Qy   2161 cttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagata 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATA 2220

Qy   2221 atcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagt 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 ATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGT 2280

Qy   2281 gctgggcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||||||
Db   2281 GCTGGGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 2320


RESULT   4
AF462267
LOCUS       AF462267                2363 bp    mRNA    linear   PRI 07-FEB-2002
DEFINITION  Homo sapiens UDP-glucuronosyltransferase UGT1A8*2 (UGT1A8) mRNA,
            UGT1A8*2 allele, complete cds.
ACCESSION   AF462267
VERSION     AF462267.1  GI:18568216
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2363)
  AUTHORS   Huang,Y.-H., Geske,D., Nguyen,N. and Tukey,R.
  TITLE     Identification and functional characterization of UGT1A8*1,
            UGT1A8*2 and UGT1A8*3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2363)
  AUTHORS   Huang,Y.-H., Geske,D., Nguyen,N. and Tukey,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2001) Pharmacology, UCSD, 9500 Gilman Drive, La
            Jolla, CA 92093, USA
FEATURES             Location/Qualifiers
     source          1. .2363
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="colon"
     gene            1. .2363
                     /gene="UGT1A8"
                     /allele="UGT1A8*2"
     CDS             34. .1626
                     /gene="UGT1A8"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase UGT1A8*2"
                     /protein_id="AAL75963.1"
                     /db_xref="GI:18568217"
                     /translation="MARTGWTSPIPLCVSLLLTCGFAEAGKLLVVPMDGSHWFTMQSV
                     VEKLILRGHEVVVVMPEVSWQLGKSLNCTVKTYSTSYTLEDLDREFMDFADAQWKAQV
                     RSLFSLFLSSSNGFFNLFFSHCRSLFNDRKLVEYLKESSFDAVFLDPFDACGLIVAKY
                     FSLPSVVFARGIGCHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHL
                     FCQYFSKNALEIASEILQTPVTAYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT      617 a    503 c    551 g    692 t
ORIGIN


  Query Match           95.6%;  Score 2218.4;  DB 9;  Length 2363;
  Best Local Similarity  97.4%;  Pred. No. 0;
  Matches 2255;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy       5 gctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtgtgtc 64
          |||||||||||| |||||| |||||||||||||||||| |||| |||||||| | ||
Db      20 GCTGCAGTTCTCTCATGGCTCGCACAGGGTGGACCAGCCCCATTCCCCTATGTGTTTCTC 79

Qy      65 tgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatggga 124
           |||||||||||||||||||| ||||||||||||| ||||||||||||||||||||||||
Db      80 TGCTGCTGACCTGTGGCTTTGCTGAGGCAGGGAAGCTGCTGGTAGTGCCCATGGATGGGA 139

Qy     125 gccactggttcaccatgaggtcggtggtggagaaactcattctcagggggcatgaggtgg 184
           | ||||||||||||||  ||||||||||||||||||| || ||||||||||||||||||||
Db     140 GTCACTGGTTCACCATGCAGTCGGTGGTGGAGAAACTTATCCTCAGGGGGCATGAGGTGG 199

Qy     185 ttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaaaga 244
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     200 TTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGGAAAATCACTGAATTGCACAGTGAAGA 259

Qy     245 cttattcaacttcatataccctggaggatctggaccgggagttcaaggctttgcccatg 304
           |||| ||||| ||||| || ||||||||||||||||||||| |||| || ||| ||| |||
Db     260 CTTACTCAACCTCATACACTCTGGAGGATCTGGACCGGGAATTCATGGATTTCGCCGATG 319

Qy     305 ctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgaca 364
           |||||||||||||||||||||||||||| | | ||||||||| || ||||| |||||
Db     320 CTCAATGGAAAGCACAAGTACGAAGTTTGTTTTCTCTATTTCTGAGTTCATCCAATGGTT 379

Qy     365 tttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaat 424
           ||||| |||||||||||||  |||||||||||||||||||| |||  |||||||||||||
Db     380 TTTTTAACTTATTTTTTTCGCATTGCAGGAGTTTGTTTAATGACCGAAAATTAGTAGAAT 439

Qy     425 acttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggcttaa 484
           |||||||||||||||||||||||| |||||||| ||||||||||  ||||||||||||||
Db     440 ACTTAAAGGAGAGTTCTTTTGATGCGGTGTTTCTTGATCCTTTTGATGCCTGTGGCTTAA 499

Qy     485 ttgttgccaaatatttctccctccccctccgtggtcttcgccaggggaatactttgccact 544
           |||||||||||||||||||||||||||||| ||||||||||||||||||||  ||||||||
Db     500 TTGTTGCCAAATATTTCTCCCTCCCCTCTGTGGTCTTCGCCAGGGGAATAGGTTGCCACT 559

Qy     545 atcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttag 604
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     560 ATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTCTTAG 619

Qy     605 ggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggagg 664
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     620 GGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTACGGAACCACATCATGCACTTGGAGG 679

Qy     665 aacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaattctcc 724
           |||||||||| |||||   ||||| ||||||||||||||||||||||||||||||||||||
Db     680 AACATTTATTTTGCCAGTATTTTTCCAAAAATGCCCTAGAAATAGCCTCTGAAATTCTCC 739

Qy     725 aaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacgg 784
           |||||||||| || |  ||||||||||||||||||||||||||||||||||||||||||| |
Db     740 AAACACCTGTCACAGCATATGATCTCTACAGCCACACATCAATTTGGTTGTTGCGAACAG 799

Qy     785 actttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatca 844
           |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db     800 ACTTTGTTTTGGACTATCCCAAACCCGTGATGCCCAATATGATCTTCATTGGTGGTATCA 859

Qy     845 actgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctggag 904
           ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db     860 ACTGCCATCAGGGAAAGCCATTGCCTATGGAATTTGAAGCCTACATTAATGCTTCTGGAG 919

Qy     905 aacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaagaaag 964
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     920 AACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAG 979

Qy     965 ctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactg 1024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     980 CTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTG 1039

Qy    1025 gaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacg 1084
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1040 GAACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACG 1099

Qy    1085 atctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgttt 1144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1100 ATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTT 1159

Qy    1145 atgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatgg 1204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1160 ATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGG 1219
```

```
  Matches 2320;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTCAGCTGCAGTTCTCTGATGGCTTGCACAGGGTGGACCAGCCCCCTTCCTCTATGTGTG 60

Qy      61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGTCTGCTGCTGACCTGTGGCTTTGCCGAGGCAGGGAAGCTACTGGTAGTGCCCATGGAT 120

Qy     121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggggcatgag 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGGAGCCACTGGTTCACCATGAGGTCGGTGGTGGAGAAACTCATTCTCAGGGGGCATGAG 180

Qy     181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GTGGTTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGGAAGATCACTGAATTGCACAGTG 240

Qy     241 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AAGACTTATTCAACTTCATATACCCTGGAGGATCTGGACCGGGAGTTCAAGGCTTTTGCC 300

Qy     301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CATGCTCAATGGAAAGCACAAGTACGAAGTATATATTCTCTATTAATGGGTTCATACAAT 360

Qy     361 gacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagta 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GACATTTTTGACTTATTTTTTTCAAATTGCAGGAGTTTGTTTAAAGACAAAAAATTAGTA 420

Qy     421 gaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GAATACTTAAAGGAGAGTTCTTTTGATGCAGTGTTTCTCGATCCTTTTGATAACTGTGGC 480

Qy     481 ttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TTAATTGTTGCCAAATATTTCTCCCTCCCCTCCGTGGTCTTCGCCAGGGGAATACTTTGC 540

Qy     541 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CACTATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTC 600

Qy     601 ttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TTAGGGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTACGGAACCACATCATGCACTTG 660

Qy     661 gaggaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaatt 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GAGGAACATTTATTATGCCACCGTTTTTTCAAAAATGCCCTAGAAATAGCCTCTGAAATT 720

Qy     721 ctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcga 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CTCCAAACACCTGTTACGGAGTATGATCTCTACAGCCACACATCAATTTGGTTGTTGCGA 780

Qy     781 acggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggt 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 ACGGACTTTGTTTTGGACTATCCCAAACCCGTGATGCCCAACATGATCTTCATTGGTGGT 840

Qy     841 atcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttct 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ATCAACTGCCATCAGGGAAAGCCGTTGCCTATGGAATTTGAAGCCTACATTAATGCTTCT 900

Qy     901 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GGAGAACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAG 960

Qy     961 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 AAAGCTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTAC 1020

Qy    1021 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaa 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 ACTGGAACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAA 1080

Qy    1081 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 AACGATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGT 1140

Qy    1141 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAG 1200

Qy    1201 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ATGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAA 1260

Qy    1261 atgacttctgaagatttagaaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 ATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAG 1320

Qy    1321 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 GAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTG 1380

Qy    1381 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 GCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCA 1440

Qy    1441 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 GCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCC 1500

Qy    1501 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 GTCGTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGC 1560

Qy    1561 ttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggt 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 TTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGT 1620

Qy    1621 gggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatc 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 GGGAAATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATC 1680

Qy    1681 agtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccag 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 AGTGTTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAG 1740

Qy    1741 agtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatattt 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 AGTGCTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTT 1800

Qy    1801 gaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcaga 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 GAATATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGA 1860

Qy    1861 ggacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgttt 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 GGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTT 1920

Qy    1921 gggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaat 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 GGGAGTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAAT 1980

Qy    1981 cttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcct 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 CTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCT 2040

Qy    2041 catctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacc 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 CATCTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACC 2100

Qy    2101 tttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgattt 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 TTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTT 2160
```

```
Db   1021 ACTGGAACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAA 1080

Qy   1081 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 AACGATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGT 1140

Qy   1141 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAG 1200

Qy   1201 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ATGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAA 1260

Qy   1261 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 ATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAG 1320

Qy   1321 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 GAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTG 1380

Qy   1381 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCA 1440

Qy   1441 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCC 1500

Qy   1501 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GTCGTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGC 1560

Qy   1561 ttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggt 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGT 1620

Qy   1621 gggaaataaggtaaaatttttgaaccattccctagtcatttccaaacttgaaaacagaatc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GGGAAATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATC 1680

Qy   1681 agtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccag 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 AGTGTTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAG 1740

Qy   1741 agtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatattt 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AGTGCTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTT 1800

Qy   1801 gaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcaga 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 GAATATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGA 1860

Qy   1861 ggacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgttt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTT 1920

Qy   1921 gggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaat 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GGGAGTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAAT 1980

Qy   1981 cttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcct 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCT 2040

Qy   2041 catctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacc 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CATCTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACC 2100

Qy   2101 tttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgattt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 TTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTT 2160
```

```
Qy   2161 cttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagata 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATA 2220

Qy   2221 atcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagt 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 ATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGT 2280

Qy   2281 gctgggcaagtttactttttttctgatgtttcctacaact 2320
          ||||||||||||||||||||||||||||||||||||||||
Db   2281 GCTGGGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 2320
```

```
RESULT   3
AF056188
LOCUS       AF056188                2320 bp    mRNA    linear   PRI 13-AUG-1998
DEFINITION  Homo sapiens UDP-glucuronosyltransferase 1A9 (UGT1A9) mRNA,
            complete cds.
ACCESSION   AF056188
VERSION     AF056188.1  GI:3414797
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2320)
  AUTHORS   Ciotti,M., Potter,C. and Owens,I.S.
  TITLE     Human phenol metabolizing UDP-glucuronosyltransferase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2320)
  AUTHORS   Ciotti,M., Potter,C. and Owens,I.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-1998) Heritable Disorders Branch, NICHD, 9000
            Rockville Pike, Bethesda, MD 20892, USA
REFERENCE   3  (bases 1 to 2320)
  AUTHORS   Ciotti,M., Potter,C. and Owens,I.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-AUG-1998) Heritable Disorders Branch, NICHD, 9000
            Rockville Pike, Bethesda, MD 20892, USA
  REMARK    Sequence update by submitter
COMMENT     On Aug 13, 1998 this sequence version replaced gi:3025895.
FEATURES             Location/Qualifiers
     source          1. .2320
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
                     /tissue_type="liver"
     gene            1. .2320
                     /gene="UGT1A9"
     CDS             19. .1611
                     /gene="UGT1A9"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A9"
                     /protein_id="AAC31425.1"
                     /db_xref="GI:3025896"
                     /translation="MACTGWTSPLPLCVCLLLTCGFAEAGKLLVVPMDGSHWFTMRSV
                     VEKLILRGHEVVVVMPEVSWQLGRSLNCTVKTYSTSYTLEDLDREFKAFAHAQWKAQV
                     RSIYSLLMGSYNDIFDLFFSNCRSLFKDKKLVEYLKESSFDAVFLDPFDNCGLIVAKY
                     FSLPSVVFARGILCHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHL
                     LCHRFFKNALEIASEILQTPVTEYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT      608 a    495 c    547 g    670 t
ORIGIN
```

```
  Query Match           100.0%;  Score 2320;  DB 9;  Length 2320;
  Best Local Similarity 100.0%;  Pred. No. 0;
```

```
Db   1861 GGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTT 1920

Qy   1921 gggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaat 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GGGAGTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAAT 1980

Qy   1981 cttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcct 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCT 2040

Qy   2041 catctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacc 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CATCTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACC 2100

Qy   2101 tttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgattt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 TTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTT 2160

Qy   2161 cttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagata 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATA 2220

Qy   2221 atcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagt 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 ATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGT 2280

Qy   2281 gctgggcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||||||
Db   2281 GCTGGGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 2320


RESULT   2
AX286764
LOCUS       AX286764                2320 bp    DNA     linear   PAT 21-NOV-2001
DEFINITION  Sequence 3 from Patent WO0180896.
ACCESSION   AX286764
VERSION     AX286764.1  GI:17048799
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ratain,M.J., Innocenti,F. and Iyer,L.
  TITLE     Flavopiridol drug combinations and methods with reduced side
            effects
  JOURNAL   Patent: WO 0180896-A 3 01-NOV-2001;
            ARCH DEVELOPMENT CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2320
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             19. .1611
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD12275.1"
                     /db_xref="GI:17048800"
                     /translation="MACTGWTSPLPLCVCLLLTCGFAEAGKLLVVPMDGSHWFTMRSV
                     VEKLILRGHEVVVVMPEVSWQLGRSLNCTVKTYSTSYTLEDLDREFKAFAHAQWKAQV
                     RSIYSLLMGSYNDIFDLFFSNCRSLFKDKKLVEYLKESSFDAVFLDPFDNCGLIVAKY
                     FSLPSVVFARGILCHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHL
                     LCHRFFKNALEIASEILQTPVTEYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT      608 a    495 c    547 g    670 t
ORIGIN


  Query Match             100.0%;  Score 2320;  DB 6;  Length 2320;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2320;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTCAGCTGCAGTTCTCTGATGGCTTGCACAGGGTGGACCAGCCCCCTTCCTCTATGTGTG 60

Qy     61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGTCTGCTGCTGACCTGTGGCTTTGCCGAGGCAGGGAAGCTACTGGTAGTGCCCATGGAT 120

Qy    121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggggcatgag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGGAGCCACTGGTTCACCATGAGGTCGGTGGTGGAGAAACTCATTCTCAGGGGGCATGAG 180

Qy    181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTGGTTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGGAAGATCACTGAATTGCACAGTG 240

Qy    241 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAGACTTATTCAACTTCATATACCCTGGAGGATCTGGACCGGGAGTTCAAGGCTTTTGCC 300

Qy    301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CATGCTCAATGGAAAGCACAAGTACGAAGTATATATTCTCTATTAATGGGTTCATACAAT 360

Qy    361 gacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagta 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GACATTTTTGACTTATTTTTTTCAAATTGCAGGAGTTTGTTTAAAGACAAAAAATTAGTA 420

Qy    421 gaatacttaaaggagagttctttttgatgcagtgtttctcgatccttttgataactgtggc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GAATACTTAAAGGAGAGTTCTTTTGATGCAGTGTTTCTCGATCCTTTTGATAACTGTGGC 480

Qy    481 ttaattgttgccaaatatttctccctccccctccgtggtcttcgccaggggaatactttgc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTAATTGTTGCCAAATATTTCTCCCTCCCCTCCGTGGTCTTCGCCAGGGGAATACTTTGC 540

Qy    541 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CACTATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTC 600

Qy    601 ttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTAGGGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTACGGAACCACATCATGCACTTG 660

Qy    661 gaggaacatttattatgccaccgtttttttcaaaaatgccctagaaatagcctctgaaatt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GAGGAACATTTATTATGCCACCGTTTTTTCAAAAATGCCCTAGAAATAGCCTCTGAAATT 720

Qy    721 ctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcga 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CTCCAAACACCTGTTACGGAGTATGATCTCTACAGCCACACATCAATTTGGTTGTTGCGA 780

Qy    781 acggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggt 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ACGGACTTTGTTTTGGACTATCCCAAACCCGTGATGCCCAACATGATCTTCATTGGTGGT 840

Qy    841 atcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttct 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ATCAACTGCCATCAGGGAAAGCCGTTGCCTATGGAATTTGAAGCCTACATTAATGCTTCT 900

Qy    901 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GGAGAACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAG 960

Qy    961 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AAAGCTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTAC 1020

Qy   1021 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
                     /codon_start=1
                     /protein_id="CAD12274.1"
                     /db_xref="GI:17048798"
                     /translation="MACTGWTSPLPLCVCLLLTCGFAEAGKLLVVPMDGSHWFTMRSV
                     VEKLILRGHEVVVVMPEVSWQLGRSLNCTVKTYSTSYTLEDLDREFKAFAHAQWKAQV
                     RSIYSLLMGSYNDIFDLFFSNCRSLFKDKKLVEYLKESSFDAVFLDPFDNCGLIVAKY
                     FSLPSVVFARGILCHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHL
                     LCHRFFKNALEIASEILQTPVTEYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
BASE COUNT      608 a    495 c    547 g    670 t
ORIGIN
```

Query Match 100.0%; Score 2320; DB 6; Length 2320;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60
Db       1 CTCAGCTGCAGTTCTCTGATGGCTTGCACAGGGTGGACCAGCCCCCTTCCTCTATGTGTG 60

Qy      61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120
Db      61 TGTCTGCTGCTGACCTGTGGCTTTGCCGAGGCAGGGAAGCTACTGGTAGTGCCCATGGAT 120

Qy     121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcagggggcatgag 180
Db     121 GGGAGCCACTGGTTCACCATGAGGTCGGTGGTGGAGAAACTCATTCTCAGGGGGCATGAG 180

Qy     181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240
Db     181 GTGGTTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGGAAGATCACTGAATTGCACAGTG 240

Qy     241 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 300
Db     241 AAGACTTATTCAACTTCATATACCCTGGAGGATCTGGACCGGGAGTTCAAGGCTTTTGCC 300

Qy     301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360
Db     301 CATGCTCAATGGAAAGCACAAGTACGAAGTATATATTCTCTATTAATGGGTTCATACAAT 360

Qy     361 gacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagta 420
Db     361 GACATTTTTGACTTATTTTTTTCAAATTGCAGGAGTTTGTTTAAAGACAAAAAATTAGTA 420

Qy     421 gaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggc 480
Db     421 GAATACTTAAAGGAGAGTTCTTTTGATGCAGTGTTTCTCGATCCTTTTGATAACTGTGGC 480

Qy     481 ttaattgttgccaaatatttctccctcccctccgtggtcttcgccagggaatactttgc 540
Db     481 TTAATTGTTGCCAAATATTTCTCCCTCCCCTCCGTGGTCTTCGCCAGGGGAATACTTTGC 540

Qy     541 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 600
Db     541 CACTATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTC 600

Qy     601 ttagggttctcagatgccatgacttcaaggagagagtacggaaccacatcatgcacttg 660
Db     601 TTAGGGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTACGGAACCACATCATGCACTTG 660

Qy     661 gaggaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaatt 720
Db     661 GAGGAACATTTATTATGCCACCGTTTTTTCAAAAATGCCCTAGAAATAGCCTCTGAAATT 720

Qy     721 ctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcga 780
Db     721 CTCCAAACACCTGTTACGGAGTATGATCTCTACAGCCACACATCAATTTGGTTGTTGCGA 780

Qy     781 acggacttttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggt 840
Db     781 ACGGACTTTGTTTTGGACTATCCCAAACCCGTGATGCCCAACATGATCTTCATTGGTGGT 840

Qy     841 atcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttct 900
Db     841 ATCAACTGCCATCAGGGAAAGCCGTTGCCTATGGAATTTGAAGCCTACATTAATGCTTCT 900

Qy     901 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 960
Db     901 GGAGAACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAG 960

Qy     961 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1020
Db     961 AAAGCTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTAC 1020

Qy    1021 actggaaccccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaa 1080
Db    1021 ACTGGAACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAA 1080

Qy    1081 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1140
Db    1081 AACGATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGT 1140

Qy    1141 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1200
Db    1141 GTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAG 1200

Qy    1201 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1260
Db    1201 ATGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAA 1260

Qy    1261 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1320
Db    1261 ATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAG 1320

Qy    1321 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1380
Db    1321 GAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTG 1380

Qy    1381 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1440
Db    1381 GCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCA 1440

Qy    1441 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1500
Db    1441 GCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCC 1500

Qy    1501 gtcgtgctgacagtggccttcatcaccttaaatgttgtgcttatggctaccggaaatgc 1560
Db    1501 GTCGTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGC 1560

Qy    1561 ttggggaaaaaaggcgagttaagaaagcccacaaatccaagacccattgagaagtgggt 1620
Db    1561 TTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGT 1620

Qy    1621 gggaaataaggtaaaatttgaaccattccctagtcatttccaaacttgaaaacagaatc 1680
Db    1621 GGGAAATAAGGTAAAATTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATC 1680

Qy    1681 agtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccag 1740
Db    1681 AGTGTTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAG 1740

Qy    1741 agtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatattt 1800
Db    1741 AGTGCTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTT 1800

Qy    1801 gaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcaga 1860
Db    1801 GAATATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGA 1860

Qy    1861 ggacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgttt 1920
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 21, 2002, 02:50:17 ; Search time 8248.67 Seconds
(without alignments)
5885.745 Million cell updates/sec

Title: US-09-835-082-1
Perfect score: 2320
Sequence: 1 ctcagctgcagttctctgat.....ttctgatgtttcctacaact 2320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2320	100.0	2320	6	AX286762	AX286762 Sequence
2	2320	100.0	2320	6	AX286764	AX286764 Sequence
3	2320	100.0	2320	9	AF056188	AF056188 Homo sapi
4	2218.4	95.6	2363	9	AF462267	AF462267 Homo sapi
5	2207.4	95.1	2394	9	AF462268	AF462268 Homo sapi
6	2194.6	94.6	2423	9	BC020971	BC020971 Homo sapi
7	2156.8	93.0	2392	9	S55985	S55985 Homo sapien
8	2076	89.5	2396	9	AF104336	AF104336 Macaca fa
9	1895.8	81.7	2003	9	AK025403	AK025403 Homo sapi
10	1636.6	70.5	2368	9	HUMHUGBR2	M57951 Human bilir
11	1629	70.2	2388	9	BC019861	BC019861 Homo sapi
12	1610.8	69.4	2351	9	HUMHUGBR1	M57899 Human bilir
13	1539.2	66.3	2422	6	AX336711	AX336711 Sequence
14	1539.2	66.3	2422	9	HUMUGT1FA	J04093 Homo sapien
15	1493.8	64.4	1593	9	AF030310	AF030310 Homo sapi
16	1484.2	64.0	1593	9	HSU89507	U89507 Human UDP-g
17	1477	63.7	1621	9	AF104338	AF104338 Macaca fa
18	1470.6	63.4	2440	9	AF104337	AF104337 Macaca fa
19	1463.6	63.1	1593	9	HSU89508	U89508 Human UDP-g
20	1157.8	49.9	2241	10	AF461735	AF461735 Rattus no
21	1129.4	48.7	2203	10	BC019434	BC019434 Mus muscu
22	1123.4	48.4	1590	4	AF011902	AF011902 Oryctolag
23	1110.4	47.9	2155	10	MUSUGTP	L27122 Mus musculu
24	1105.2	47.6	2332	4	AB018478	AB018478 Ovis arie
25	1105	47.6	2282	10	AF461738	AF461738 Rattus no
26	1004.4	43.3	1561	9	HUMUGT03	M84124 Human UDP-g
c 27	1004.4	43.3	68770	9	AC006985	AC006985 Homo sapi
c 28	1004.4	43.3	176619	2	AC026497	AC026497 Homo sapi
29	1004.4	43.3	198872	9	AF297093	AF297093 Homo sapi
30	937	40.4	2301	10	RNU75903	U75903 Rattus norv
31	878.6	37.9	1255	9	AF297091	AF297091 Homo sapi
c 32	878.6	37.9	148438	9	AC019072	AC019072 Homo sapi
33	873	37.6	1602	9	AF360121	AF360121 Macaca mu
34	872.2	37.6	1647	9	AF104339	AF104339 Macaca fa
35	868.2	37.4	1602	9	AY029169	AY029169 Macaca mu
36	819.8	35.3	2816	4	AF039138	AF039138 Felis cat
c 37	799.8	34.5	183826	2	AC091778	AC091778 Papio cyn
38	794	34.2	2300	10	MMU16818	U16818 Mus musculu
39	788.4	34.0	2037	4	OCU09101	U09101 Oryctolagus
40	783.2	33.8	1765	10	RATUDPGTA	M34007 Rat bilirub
41	781.8	33.7	2343	10	AF461737	AF461737 Rattus no
42	779	33.6	1244	9	HSU42604	U42604 Human UDP-g
43	777.8	33.5	1927	10	RATUDPGT	J02612 Rat UDP-glu
44	774	33.4	2239	10	RNU20551	U20551 Rattus norv
45	772.2	33.3	2484	10	RATUGT	J05132 Rat 3-methy

ALIGNMENTS

RESULT 1
AX286762
LOCUS AX286762 2320 bp DNA linear PAT 21-NOV-2001
DEFINITION Sequence 1 from Patent WO0180896.
ACCESSION AX286762
VERSION AX286762.1 GI:17048797
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Ratain,M.J., Innocenti,F. and Iyer,L.
TITLE Flavopiridol drug combinations and methods with reduced side effects
JOURNAL Patent: WO 0180896-A 1 01-NOV-2001;
ARCH DEVELOPMENT CORPORATION (US)
FEATURES Location/Qualifiers
source 1. .2320
/organism="Homo sapiens"
/db_xref="taxon:9606"
CDS 19. .1611
/note="unnamed protein product"

```
Qy    544 tatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctctta 603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||    |||||||
Db    558 CATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAATGATCTCTTA 617

Qy    604 gggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggag 663
          |||||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||
Db    618 GGGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTATGGAACCACATCGTGCACTTGGAG 677

Qy    664 gaacatttattatgccaccgtttttttcaaaaatgccctagaaatagcctctgaaattctc 723
          || |||||||| |||||   | || | | ||||||||||||||||| ||||||||||||
Db    678 GACCATTTATT-TGCCAGTATCTTATTAGAAATGCCCTAGAAATAG-CTCTGAAATTCTC 735

Qy    724 caaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacg 783
          ||||| ||||| |||| ||||||||||||||| ||| ||||||||||||||||||||||
Db    736 CAAACCCCTGTCACGGCATATGATCTCTACAGTCACCCATCAATTTGGTTGTTGCGAACG 795

Qy    784 gactttgttttggac-tatcccaaacccgtgatg 816
          | | |||||| |||| ||| || |||||||| ||
Db    796 GGCCTTGTTTGGGACTTATTCCGAACCCGTGCTG 829


RESULT 15
BG400591
LOCUS       BG400591                833 bp    mRNA    linear   EST 12-MAR-2001
DEFINITION  602464244F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4592315 5',
            mRNA sequence.
ACCESSION   BG400591
VERSION     BG400591.1  GI:13294039
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 833)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1330  row: h  column: 12
            High quality sequence stop: 800.
FEATURES             Location/Qualifiers
     source          1. .833
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4592315"
                     /clone_lib="NIH_MGC_75"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.65
                     kb (range 0.5-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      206 a    179 c    211 g    237 t
ORIGIN


  Query Match             27.6%;  Score 640;  DB 10;  Length 833;
  Best Local Similarity   91.1%;  Pred. No. 1.3e-156;
  Matches  760,  Conservative    0;  Mismatches   60;  Indels   14;  Gaps    7;

Qy      8 gcagttctctgatggcttgcacagggtggaccagcccccttcctctatgtgtgtgtctgc 67
          | ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGAGTTCTCTGATGGCTTGCACAGGGTGGACCAGCCCCCTTCCTCTATGTGTGTGTCTGC 60

Qy     68 tg--ctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatgggag 125
          ||   |||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGTCGTGACCTGTGGCTTTGCCGAGGCAGGGAAGCTACTGGTAGTGCCCATGGATGGGAG 120

Qy    126 ccactggttcaccatgaggtcggtggtggagaaactcattctcaggggggcatgaggtggt 185
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCACTGGTTCACCATGAGGTCGGTGGTGGAGAAACTCATTCTCAGGGGGCATGAGGTGGT 180

Qy    186 tgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaagac 245
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGTAGTCATGCCAGAGGTGAGTTGGCAACTGGGAAGATCACTGAATTGCACAGTGAAGAC 240

Qy    246 ttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcccatgc 305
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTATTCAACTTCATATACCCTGGAGGATCTGGACCGGGAGTTCAAGGCTTTTGCCCATGC 300

Qy    306 tcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgacat 365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCAATGGAAAGCACAAGTACGAAGTATATATTCTCTATTAATGGGTTCATACAATGAGCA 360

Qy    366 ttttg----acttatttttttcaaattgcaggagtttgtttaaagacaaa-aaattagta 420
          |||||       | ||| ||  || |||||||||||||||||||||| |||||||||
Db    361 TTTTGTGACTTATTTTTGTTCAAAGTTGCAGGAGTTTGTTTAAAGACAAAGAAATTAGTA 420

Qy    421 gaatact-taaaggagagttcttttgatgcagtg---tttctcgatccttttgataactg 476
          ||||||| |||||||||||||| ||||||||||    | ||  |  | |||    |  |
Db    421 GAATACTGTAAAGGAGAGTTCTGTTGATGCAGTGTGATCTCGAGTCCGTTTGATAACTGG 480

Qy    477 tggcttaatggttgccaaatatttctccctcccctccgtggtcttcgccaggggaatact 536
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGGCTTAATGGTTGCCAAATATTTCTCCCTCCCCTCCGTGGTCTTCGCCAGGGGAATACT 540

Qy    537 ttgccactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaat 596
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTGCCACTATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAAT 600

Qy    597 tctcttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgca 656
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TCTCTTAGGGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTACGGAACCACATCATGCA 660

Qy    657 cttgg--aggaacatttattatgccaccgtttttttcaaaaatgccctagaaatagcctct 714
          |||||  |||||||| |||||||||||||| || |  |||      |    ||    ||||
Db    661 CTTGGGAAGGAACATTGATTATGCCACCGTGTTGTTCAAAGAATGCCTAGAAATAGCTCT 720

Qy    715 gaaattctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttg 774
          |||||||||||||||||||||||||||||||||||||    |      |  | |||||||
Db    721 GAAATTCTCCAAACACCTGTTACGGAGTATGATCTCTAACAGCCACACTTCAATTGGTTG 780

Qy    775 ttgcgaacggactttgttttggactatcccaaacccgtgatgcccaacatgatc 828
          |||||||||||||||| ||||||||||||||||||  ||||||||||||||||
Db    781 TTGCGAACGGACTTTG-TTTGGACTATCCCAAACCCGGAATGCCCAACATGATC 833
```

Search completed: August 21, 2002, 09:39:10
Job time: 24958 sec

```
ORIGIN


  Query Match           27.8%;  Score 644;  DB 10;  Length 723;
  Best Local Similarity  96.7%;  Pred. No. 1.1e-157;
  Matches  700;  Conservative    0;  Mismatches   20;  Indels    4;  Gaps    4;

Qy   1078 caaaacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccat 1137
          ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db      2 CAAAACGATCTGCTTGGTCACCCGATGACCCGTG-CTTTATCACCCATGCTGGTTCCCAT 60

Qy   1138 ggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgat 1197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGTGTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGAT 120

Qy   1198 cagatggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctg 1257
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CAGATGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTG 180

Qy   1258 gaaatgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttac 1317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GAAATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTAC 240

Qy   1318 aaggagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggac 1377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAGGAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGAC 300

Qy   1378 ctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgcccc 1437
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCC 360

Qy   1438 gcagcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttg 1497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCAGCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTG 420

Qy   1498 gccgtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaa 1557
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GCCGTCTTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAA 480

Qy   1558 tgcttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtg 1617
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGCTTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTG 540

Qy   1618 ggtgggaaataaggtaaaa-ttttgaaccattccctagtcatttccaaacttgaaaacag 1676
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    541 GGTGGGAAATAAGGTAAAATTTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAG 600

Qy   1677 aatcagtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagcc 1736
          |||||||||||| |||||||||||||||||||||||||| |||||||||||||||| | 
Db    601 AATCAGTGTTACATTCATTTTATTCTTATTAAGGAAATAC-TTGCATAAATTAATCAG-C 658

Qy   1737 ccagagtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagat 1796
          |||||||||||||||||||||||||  ||| |||      | |  ||| ||    |||
Db    659 CCAGAGTGCTTTAAAAAATTCTCTTAATAAAATATAGACTCGTAGTCAGTAAGATTTGAT 718

Qy   1797 attt 1800
          |||| 
Db    719 ATGT 722


RESULT  14
BG469506
LOCUS       BG469506                 829 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602532818F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4660540 5',
            mRNA sequence.
ACCESSION   BG469506
VERSION     BG469506.1  GI:13401781
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 829)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1458  row: c  column: 05
            High quality sequence stop: 795.
FEATURES             Location/Qualifiers
     source          1. .829
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4660540"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      186 a    190 c    207 g    246 t
ORIGIN


  Query Match           27.6%;  Score 641.2;  DB 10;  Length 829;
  Best Local Similarity  89.9%;  Pred. No. 6.4e-157;
  Matches  732;  Conservative    0;  Mismatches   78;  Indels    4;  Gaps    4;

Qy        5 gctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtgtgtc 64
            |||||||||||||| |||||| || |||||||||||||||||| ||||| ||||||||||||
Db       18 GCTGCAGTTCTCTCATGGCTCGCGCAGGGTGGACCAGCCCCGTTCCTTTATGTGTGTGTC 77

Qy       65 tgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatggga 124
            | |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       78 TACTGCTGACCTGTGGCTTTGCCGAGGCAGGGAAGCTGCTGGTAGTGCCCATGGATGGGA 137

Qy      125 gccactggttcaccatgaggtcggtggtggagaaactcattctcaggggggcatgaggtgg 184
            | |||||||||||||||  |||||||||||||||||| || ||||||||||||||||||||
Db      138 GTCACTGGTTCACCATGCAGTCGGTGGTGGAGAAACTTATCCTCAGGGGGCATGAGGTGG 197

Qy      185 ttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaaga 244
            |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      198 TTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGAAAGATCACTGAATTGCACAGTGAAGA 257

Qy      245 cttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcccatg 304
            |||| ||||| || || || ||||| |||| | ||||||| |||| || ||| |||||||
Db      258 CTTACTCAACCTCGTACACTCTGGAAGATCAGAACCGGGAATTCATGGTTTTCGCCCATG 317

Qy      305 ctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgaca 364
            |||||||||||||||  | || |||||||| |||||||||||| |||||| || ||
Db      318 CTCAATGGAAAGCACAGGCACAAAGTATATTTTCTCTATTAATGAGTTCATCCAGTGGTT 377

Qy      365 tttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaat 424
            || |||||||||||||||  |||||||||||| |||||| |||  ||||||||||||||
Db      378 TTCTTGACTTATTTTTTTCGCATTGCAGGAGTTGGTTTAATGACCGAAAATTAGTAGAAT 437

Qy      425 acttaaaggagagttcttttgatgcagtgtttc-tcgatccttttgataactgtggctta 483
            |||||||||||||||||  |||||||||||||| | ||||||||||| ||||||||||
Db      438 ACTTAAAGGAGAGTTCTTTGGATGCAGTGTTTCGTGGATCCTTTTGATACCTGTGGCTTA 497

Qy      484 attgttgccaaatatttctccctccccctccgtggtcttcgccaggggaatactttgccac 543
            ||||| || |||||||||||||||||||| ||||||||| |||||||||| || |||||
Db      498 ATTGTCGCTAAATATTTCTCCCTCCCCTCTGTGGTCTTCACCAGGGGAATATTTGGCCAC 557
```

```
ACCESSION   AI831562
VERSION     AI831562.1  GI:5452233
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteieostomi;
            Mammaiia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 764)
  AUTHORS   NCI-CGAP http://www.ncbi.nim.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpubiished (i997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP cione distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.iinl.gov/bbrp/image/image.html
            Insert Length: 1491   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quaiity sequence stop: 454.
FEATURES             Location/Quaiifiers
     source          1. .764
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2405161"
                     /clone_iib="NCI_CGAP_Lul9"
                     /tissue_type="squamous celi carcinoma, poorly
                     differentiated (4 pooled tumors, including primary and
                     metastatic)"
                     /dev_stage="adult"
                     /iab_host="DH10B (phage-resistant)"
                     /note="Organ: iung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     pooled iung tumor tissue, and was then primed with a Not I
                     - oligo(dT) primer. Double-stranded cDNA was iigated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cioned into the Not I  and Eco RI sites of the modified
                     pT7T3 vector.  Library went througb one round of
                     normalization. Library constructed by Bento Soares and M.
                     Fatima Bonaido. "
BASE COUNT      242 a    156 c    133 g    232 t      1 others
ORIGIN
```

```
  Query Match           28.9%;  Score 669.4;  DB 9;  Length 764;
  Best Local Similiarity   96.8%;  Pred. No. 2.5e-164;
  Matches  693;  Conservative    0;  Mismatches   22;  Indels    1;  Gaps    1;

Qy   1605 ccattgagaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttccaa 1664
           ||   ||  ||||||||||||||| |||||| |||||||||| ||   ||||||||||
Db    755 CCCATGAAGAGTGGGTGGGAAATAGGGTAAATTTTTGAACCATCCCTACGTCATTTCCAA 696

Qy   1665 acttgaaaacagaatcagtgttaaattcattttattcttattaaggaaatactttgcata 1724
             |||||||    |   |||||||||||||| |||||||||||||| ||||||||||||
Db    695 CCTTGAAACAGCATCAAGTGTTAAATTCATCTTATTCTTATTAAGAAAATACTTTGCATA 636

Qy   1725 aattaatcagccccagagtgctttaaaaaattctcttaaatataaaaataatagactcgcta 1784
           |||||||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||
Db    635 AATTAATCAGCCCCAGAGTGC-TTAAANAATTCTCTTAAATAAAAATAATAGACTCGCTA 577

Qy   1785 gtcagtaaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgaca 1844
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    576 GTCAGTACAGATATTTGAATATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACA 517

Qy   1845 tgtgccatttttcagaggacgtgcagacaggctggcattctagattacttttcttactct 1904
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    516 TGTGCCATTTTTCAGAGGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCT 457
```

```
Qy   1905 gaaacatggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctactgc 1964
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    456 GAAACATGGCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGC 397

Qy   1965 aaatggcagttttaatcttatcttttggcttctgcagatggttgcaattgatccttaacc 2024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    396 AAATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACC 337

Qy   2025 aataatggtcagtcctcatctctgtcctgcttcataggtgccaccttgtgtgtttaaaga 2084
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    336 AATAATGGTCAGTCCTCATCTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGA 277

Qy   2085 agggaagctttgtacctttagagtgtaggtgaaatgaatgaatggcttggagtgcactga 2144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    276 AGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGA 217

Qy   2145 gaacagcatatgatttcttgctttggggaaaaagaatgatgctatgaaattggtgggtgg 2204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    216 GAACAGCATATGATTTCTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGG 157

Qy   2205 tgtatttgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaacccaa 2264
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    156 TGTATTTGAGAAGATAATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAA 97

Qy   2265 aatacagctatgaagtgctgggcaagtttacttttttttctgatgtttcctacaact 2320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     96 AATACAGCTATGAAGTGCTGGGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 41
```

```
RESULT  13
BG424452
LOCUS       BG424452                723 bp    mRNA    linear   EST 14-MAR-2001
DEFINITION  602447772F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4586368 5',
            mRNA sequence.
ACCESSION   BG424452
VERSION     BG424452.1  GI:13330958
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteieostomi;
            Mammaiia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   i  (bases 1 to 723)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Heaith, Mammaiian Gene Coiiection (MGC)
  JOURNAL   Unpublished (i999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.iinl.gov
            Plate: LLCM1314  row: p  column: 17
            High quaiity sequence stop: 711.
FEATURES             Location/Quaiifiers
     source          1. .723
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4586368"
                     /clone_lib="NIH_MGC_14"
                     /tissue_type="renal cell adenocarcinoma"
                     /iab_host="DH10B (phage-resistant)"
                     /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oiigo-dT priming.  Directionaliy
                     cioned into EcoRI/XhoI sites using the foiiowing 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size i.8kb.  Library constructed by Ling Hong in
                     the iaboratory of Geraid M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      202 a    160 c    170 g    191 t
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGAAGATTTAGA 240

Qy   1281 aaatgctctaaaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctc 1340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAACATCATGCGCCTCTC 300

Qy   1341 cagccttcacaaggaccgcccggtggagccgctggacctggccgtgttctgggtggagtt 1400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAGTT 360

Qy   1401 tgtgatgaggcacaagggcgcgccacacctgcgccccgcagcccacgacctcacctggta 1460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCCACGACCTCACCTGGTA 420

Qy   1461 ccagtaccattccttggacgtgattggtttcctcttggccgtcgtgctgacagtggcctt 1520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGCCTT 480

Qy   1521 catcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcgagt 1580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGGAAAAAAGGGCGAGT 540

Qy   1581 taagaaagcccacaaatccaagacccattgagaagtgggtgggaaataaggtaaaatttt 1640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGAAATAAGGTAAAATTTG 600

Qy   1641 gaaccattccct-agtcatttccaaacttgaaaacagaatcagtgttaaattcatttat 1699
          |||||||||||| |||||||||||||||||||||||||||||||| ||||||| |
Db    601 GAACCATTCCCTAAGTCATTTCCAAACTTGAAAACAGAATCAGTGTTCAATTCATCT--A 658

Qy   1700 tcttattaaggaaatactttgcataaattaatcagccccagagtgctttaaaaaattctc 1759
          |||||||| |||||||||| || || ||| ||||| |||||||||||||| | |||| |||
Db    659 TCTTATTACGGAAATACTTCGCCTACATT-ATCAGGCCCAGAGTGCTTT-ACAAATCCTC 716

Qy   1760 ttaaataaaaataatagactcgctagtcagtaaagatattt 1800
          | ||  | |||||||||||||||||||||| | |||| | |
Db    717 TAAATAACAAATAATAGACTCGCTAGTCAGAACAGATCTCT 757


RESULT  11
BG747143
LOCUS       BG747143                697 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602704364F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4857831 5',
            mRNA sequence.
ACCESSION   BG747143
VERSION     BG747143.1  GI:14057796
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1710  row: o  column: 16
            High quality sequence stop: 684.
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4857831"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      177 a    168 c    184 g    168 t
ORIGIN


  Query Match             29.3%;  Score 679.2;  DB 10;  Length 697;
  Best Local Similarity   99.4%;  Pred. No. 6.5e-167;
  Matches  692;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy    910 ggaattgtggttttctctttgggatcaatggtctcagaaattccagagaagaaagctatg 969
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 GGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATG 61

Qy    970 gcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactggaacc 1029
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 GCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGGAACC 121

Qy   1030 cgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacgatctg 1089
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 CGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACGATCTG 181

Qy   1090 cttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgtttatgaa 1149
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 CTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAA 241

Qy   1150 agcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatggacaat 1209
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 AGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAAT 301

Qy   1210 gcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatgacttct 1269
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 GCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCT 361

Qy   1270 gaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggagaacatc 1329
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 GAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAACATC 421

Qy   1330 atgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccgtgttc 1389
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 ATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTC 481

Qy   1390 tgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcaggcccacgac 1449
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 TGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGGCCACGAG 541

Qy   1450 ctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcgtgctg 1509
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 CTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGTGCTG 601

Qy   1510 acagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaa 1569
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602 ACAGTGGGCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGGAAA 661

Qy   1570 aaagggcgagttaagaaagcccacaaatccaagacc 1605
          |||||||||||||||||| ||||||||||||||||
Db    662 AAAGGGCGAGTTAAGAAAG-CCACAAATCCAAGACC 696


RESULT  12
AI831562/c
LOCUS       AI831562                764 bp    mRNA    linear   EST 21-DEC-1999
DEFINITION  wj39b01.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2405161 3'
            similar to gb:M57899 UDP-GLUCURONOSYLTRANSFERASE 1A PRECURSOR,
            MICROSOMAL (HUMAN);, mRNA sequence.
```

```
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM10401  row: n  column: 18
           High quality sequence stop: 763.
FEATURES             Location/Qualifiers
     source          1. .881
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4514129"
                     /clone_lib="NIH_MGC_93"
                     /tissue_type="transitional cell papilloma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      263 a    154 c    198 g    266 t
ORIGIN


  Query Match             30.0%;  Score 695.4;  DB 10;  Length 881;
  Best Local Similarity   97.7%;  Pred. No. 4e-171;
  Matches  769;  Conservative    0;  Mismatches   11;  Indels    7;  Gaps    6;

Qy   1537 tgtgcttatggctaccggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaa 1596
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGTGCTTATGGCTACCGGAAATGCTTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAA 60

Qy   1597 tccaagacccattgagaagtgggtgggaaataaggtaaaatttttgaaccattccctagtc 1656
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCCAAGACCCATTGAGAAGTGGGTGGGAAATAAGGTAAAATTTTGAACCATTCCCTAGTC 120

Qy   1657 atttccaaacttgaaaacagaatcagtgttaaattcattttattcttattaaggaaatac 1716
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTTCCAAACTTGAAAACAGAATCAGTGTTAAATTCATTTTATTCTTATTAAGGAAATAC 180

Qy   1717 tttgcataaattaatcagccccagagtgctttaaaaaattctcttaaataaaaataatag 1776
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTTGCATAAATTAATCAGCCCCAGAGTGCTTTAAAAAATTCTCTTAAATAAAAATAATAG 240

Qy   1777 actcgctagtcagtaaagatatttgaatatgtatcgtgccccctccggtgtctttgatca 1836
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACTCGCTAGTCAGTAAAGATATTTGAATATGTATCGTGCCCCCTCCGGTGTCTTTGATCA 300

Qy   1837 ggatgacatgtgccatttttcagaggacgtgcagacaggctggcattctagattacttttt 1896
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGATGACATGTGCCATTTTTCAGAGGACGTGCAGACAGGCTGGCATTCTAGATTACTTTT 360

Qy   1897 cttactctgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcc 1956
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTTACTCTGAAACATGGCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCACCGCTGCC 420

Qy   1957 cctactgcaaatggcagttttaatcttatcttttggcttctgcagatggttgcaattgat 2016
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCTACTGCAAATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGAT 480

Qy   2017 ccttaaccaataatggtcagtcctcatctctgtcctgcttcataggtgccaccttgtgtg 2076
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CCTTAACCAATAATGGTCAGTCCTCATCTCTGTCCTGCTTCATAGGTGCCACCTTGTGTG 540

Qy   2077 tttaaagaagggaagctttgtacctttagagtgtaggtgaaatgaatgaatggcttggag 2136
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTTAAAGAAGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAG 600

Qy   2137 -tgcactgagaaca-gcatatgat-ttcttgctttggggaaaaagaatgatgctatgaaa 2193
            |||||||||||| |||||||| ||||||  ||||||||||||||||||||||||||||
Db    601 TTGCACTGAGAACAGGCATATGATCTTCTTGGCTTGGGGAAAAAGAATGATGCTATGAAA 660

Qy   2194 ttggtgggtggtgtatttgagaagataatcattgcttatgtcaaatggagctgaatttga 2253
           |||||||||||||||| ||||||||||||||| ||||||| ||||||||||||||||||
Db    661 -TGGTGGGTGGTGTATTCGAGAAGATAATCATTGTTTATGTCCAATGGAGCTGAATTTGA 719

Qy   2254 taaaaacccaaaatacagctatgaagtgctgggcaagtttacttttttttctgatgtttcc 2313
           | |||||||||||||||| ||||||||||||||  |  ||   ||||||||||||||||||
Db    720 T-AAAACCCAAAATACAGGTATGAAGTGCTGGG--AAGTTTACTTTTTTCTGATGTTTCC 776

Qy   2314 tacaact 2320
           |||||||
Db    777 TACAACT 783


RESULT  10
BE870911
LOCUS       BE870911.1               1005 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601448862F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3852831 5',
            mRNA sequence.
ACCESSION   BE870911
VERSION     BE870911.1  GI:10319687
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1005)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM9575  row: l  column: 16
           High quality sequence stop: 645.
FEATURES             Location/Qualifiers
     source          1. .1005
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3852831"
                     /clone_lib="NIH_MGC_65"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.8 kb.   Library constructed by Life
                     Technologies.  "
BASE COUNT      282 a    268 c    246 g    209 t
ORIGIN


  Query Match             29.3%;  Score 680.6;  DB 10;  Length 1005;
  Best Local Similarity   96.8%;  Pred. No. 3.2e-167;
  Matches  737;  Conservative    0;  Mismatches   19;  Indels    5;  Gaps    4;

Qy   1041 tcttgcgaacaacacgatacttgttaagtggctaccccaaaacgatctgcttggtcaccc 1100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACGATCTGCTTGGTCACCC 60

Qy   1101 gatgacccgtgcctttatcacccatgctggttcccatggtgtttatgaaagcatatgcaa 1160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAGCATATGCAA 120

Qy   1161 tggcgttcccatggtgatgatgcccttgtttggtgatcagatggacaatgcaaagcgcat 1220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAATGCAAAGCGCAT 180

Qy   1221 ggagactaagggagctggagtgaccctgaatgttctggaaatgacttctgaagatttaga 1280
```

```
Qy    683 gttttttcaaaaatgccctag--aaatagcctctgaaattctcc-aaacacctgttacgg 739
          ||||||||||||||||||||  ||    |||||||||| |||| ||||||||||||||
Db    660 GTTTTTTCAAAAATGCCCTAGGAAATAGCCCTCTGAAATNCTCCAAAACACCTGTTACGG 719

Qy    740 agtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttggact 799
          |||||||||||||||||||||||||| | || |  |    |  |   |||| |||||||
Db    720 AGTATGATCTCTACAGCCACACATCAACTCGGCTCGTTGCGAACGGACTTGTCTTGGACT 779

Qy    800 atcccaaacccgtgatgcccaacatgatcttcattggtggtatcaac--tgccatcaggg 857
          || ||||||||||||||||  |||||| || |||| |||||| ||    |    | ||
Db    780 AT-CCAAACCCGTGATGCCCACAATGATCCTCCTTGGGGGTATCCACCTGCCGTCTACGG 838

Qy    858 aaagccgttgcctatggaatttga 881
          |||||||||||||||||| || ||
Db    839 AAAGCCGTTGCCTATGGATTTGGA 862
```

RESULT 8
AI307269/c
LOCUS AI307269 818 bp mRNA linear EST 08-APR-1999
DEFINITION tb24dl2.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2055287 3' similar to gb:M57899 UDP-GLUCURONOSYLTRANSFERASE 1A PRECURSOR, MICROSOMAL (HUMAN);, mRNA sequence.
ACCESSION AI307269
VERSION AI307269.1 GI:4002458
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 818)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 453.
FEATURES Location/Qualifiers
source 1. .818
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2055287"
/clone_lib="NCI_CGAP_Kid12"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_Kid5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1323912-1325831, 1471368-1472903 and 1492104-1493255). Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 247 a 177 c 134 g 258 t 2 others
ORIGIN

Query Match 30.0%; Score 695.8; DB 9; Length 818;
Best Local Similarity 95.9%; Pred. No. 3e-171;
Matches 745; Conservative 0; Mismatches 29; Indels 3; Gaps 3;

```
Qy   1547 gctaccggaaatgcttggggaaaaaagggcg-agttaagaaagcccacaaatccaagacc 1605
          ||||||| ||||||| |||||||||||||||| | ||||| |||||| |||  |||||| |
Db    818 GCTACCGAAAATGCTGGGGGAAAAAAGGGCGAATTTAAGNAAGCCCNCAATCCCAAGAAC 759

Qy   1606 cattgagaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttccaaa 1665
            |||||||||||||||||  ||||||||||  | |  | ||| |||||||||| ||
Db    758 ATTTGAGAAGTGGGTGGGAATAAAGGTAAAATTTGAACCATTCCCCAAGTCATTTCCCAA 699

Qy   1666 cttgaaaacagaatca-gtgttaaattcattttattcttattaaggaaatactttgcata 1724
          || |||||||||||| | || ||||||||||||||||||||||||||||||||| |||||
Db    698 CTGGAAAACAGAATCAGGGGTAAAATTCATTTTATTCTTATTAAGGAAATACTTGGCATA 639

Qy   1725 aattaatcagcccc-agagtgctttaaaaaattctcttaaataaaaataatagactcgct 1783
          |||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||||
Db    638 AATTAATCAGCCCCAAGAGTGCTTTTAAAAATTCTCTTAAATAAAAATAATAGACTCGCT 579

Qy   1784 agtcagtaaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgac 1843
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    578 AGTCAGTAAAGATATTTGAATATGTATCGTGCCCCCTCTGGTGTCTTTGATCAGGATGAC 519

Qy   1844 atgtgccatttttcagaggacgtgcagacaggctggcattctagattacttttcttactc 1903
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    518 ATGTGCCATTTTTCAGAGGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTC 459

Qy   1904 tgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctactg 1963
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    458 TGAAACATGGCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCACGGCTGCCCCTACTG 399

Qy   1964 caaatggcagttttaatcttatcttttggcttctgcagatggttgcaattgatccttaac 2023
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    398 CAAATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTGGCAATTGATCCTTAAC 339

Qy   2024 caataatggtcagtcctcatctctgtcctgcttcataggtgccaccttgtgtgtttaaag 2083
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    338 CAATAATGGTCAGTCCTCATCTCTGTCGTGCTTCATAGGTGCCACCTTGTGTGTTTAAAG 279

Qy   2084 aagggaagctttgtacctttagagtgtaggtgaaatgaatgaatggcttggagtgcactg 2143
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    278 AAGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTG 219

Qy   2144 agaacagcatatgatttcttgctttggggaaaaagaatgatgctatgaaattggtgggtg 2203
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    218 AGAACAGCATATGATTTCTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTG 159

Qy   2204 gtgtatttgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaaccca 2263
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    158 GTGTATTTGAGAAGATAATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCA 99

Qy   2264 aaatacagctatgaagtgctgggcaagtttacttttttttctgatgtttcctacaact 2320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     98 AAATACAGCTATGAAGTGCTGGGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 42
```

RESULT 9
BG289777
LOCUS BG289777 881 bp mRNA linear EST 21-FEB-2001
DEFINITION 602385071F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4514129 5', mRNA sequence.
ACCESSION BG289777
VERSION BG289777.1 GI:13045907
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 881)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC

```
Qy    181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    187 GTGGTTGTAGTCATGCCAGAGGTGAGTTGGCAACTGGGAAGATCACTGAATTGCACAGTG 246

Qy    241 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    247 AAGACTTATTCAACTTCATATACCCTGGAGGATCTGGACCGGGAGTTCAAGGCTTTTGCC 306

Qy    301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    307 CATGCTCAATGGAAAGCACAAGTACGAAGTATATATTCTCTATTAATGGGTTCATACAAT 366

Qy    361 gacatttttgactta-ttttttttcaaattgcaggagtttgtttaaagacaaaaaattagt 419
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    367 GACATTTTTGACTTATTTTTTTTTCAAATTGCAGGAGTTTGTTTAAAGACAAAAAATTAGT 426

Qy    420 agaatacttaaaggagagttc-ttttgatgcagtgtttctcgatccttttgataactgtg 478
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    427 AGAATACTTAAAGGAGAGTTCTTTTTGATGCAGTGTTTCTCGATCCTTTTGATAACTGTG 486

Qy    479 gcttaattgttgccaaatatttctccctccccctccgtggtcttcgccaggggaatacttt 538
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    487 GCTTAATTGTTGCCAAATATTTCTCCCTCCCCTCCGTGGTCTTCGCCAGGCGAATACTTT 546

Qy    539 gccactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattc 598
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    547 GCCACTATCTTGAAGAAGGTGCACAGTGCCCTGCTCCTCTTTCCTATGTCCCCAGAATTC 606

Qy    599 tcttagggttctcagatgccatgactttcaaggagagagtacggaaccac-atcatgcac 657
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    607 TCTTAGGGTTCTCAGATGCCATGACTTTCAAGGAGAGAGTACGGAACCACAATCATGCAC 666

Qy    658 ttggaggaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaa 717
          |||||||||||||||||||||||||| || |||||||||||||||||||| |||||||
Db    667 TTGGAGGAACATTTATTATGCCACCGATTATTCAAAAATGCCCTAGAAATAGGCTCTGAA 726

Qy    718 attctccaaacacctgtt-acggagtatga---tctctacagccacacatcaatttggtt 773
          |||||||||| |||||||| ||||||||||   ||| | | |||||   |||| | | |
Db    727 ATTCTCCAAAGACCTGTTAACGGAGTATGAATCTCTACAGACACACACTCAAATTGGAAT 786

Qy    774 gttgcgaacgg--actttgttttggactatcccaaaccccgt-gatgcccaacatgatctt 830
          ||||||||||   ||| |||  | ||||||| ||||||||||||||||||||||||||||
Db    787 GTTGCGAACGGGACCTTGGTTGGGAACTATCCAAAACCCGTGGATGCCCAACATGATCTT 846

Qy    831 cattggtggtatcaactgccatcagggaaa 860
          ||| || |  |  |    |  || | ||||
Db    847 CATGGGAGAAACAACTGTCATTCCGTGAAA 876


RESULT   7
BI762347
LOCUS       BI762347                884 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603049004F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5189040 5',
            mRNA sequence.
ACCESSION   BI762347
VERSION     BI762347.1  GI:15753925
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 884)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM11472  row: h  column: 01
             High quality sequence stop: 796.
FEATURES             Location/Qualifiers
     source          1. .884
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5189040"
                     /clone_lib="NIH_MGC_116"
                     /lab_host="DH10B"
                     /note="Organ: pooled colon, kidney, stomach; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 3 colons, age 26 yo male, 49 yo
                     female, 71 yo male colon; 46 yo male kidney, and pool of 2
                     stomachs, 62 yo male and 70 yo female.  Library is
                     oligo-dT primed and  directionally cloned (EcoRV site is
                     destroyed upon cloning).  Average insert size 1.4 kb,
                     insert size range 1-3 kb.  Library is normalized and
                     enriched for full-length clones and was constructed by C.
                     Gruber  (Invitrogen).  Research Genetics tracking code
                     023. Note: this is a NIH_MGC Library."
BASE COUNT      213 a    207 c    210 g    253 t      1 others
ORIGIN


  Query Match           30.5%;  Score 707.4;  DB 10;  Length 884;
  Best Local Similarity   94.0%;  Pred. No. 2.8e-174;
  Matches  812;  Conservative    0;  Mismatches   42;  Indels   10;  Gaps    7;

Qy     26 gcacagggtggaccagccccttcctctatgtgtgtgtctgctgctgacctgtggctttg 85
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCACAGGGTGGACCAGCCCCCTTCCTCTATGTGTGTGTCTGCTGCTGACCTGTGGCTTTG 60

Qy     86 ccgaggcagggaagctactggtagtgcccatggatgggagccactggttcaccatgaggt 145
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCGAGGCAGGGAAGCTACTGGTAGTGCCCATGGATGGGAGCCACTGGTTCACCATGAGGT 120

Qy    146 cggtggtggagaaactcattctcaggggcatgaggtggttgtagtcatgccagaggtga 205
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGGTGGTGGAGAAACTCATTCTCAGGGGGCATGAGGTGGTTGTAGTCATGCCAGAGGTGA 180

Qy    206 gttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttcatataccc 265
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTTGGCAACTGGGAAGATCACTGAATTGCACAGTGAAGACTTATTCAACTTCATATACCC 240

Qy    266 tggaggatctggaccgggagttcaaggcttttgcccatgctcaatggaaagcacaagtac 325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGGAGGATCTGGACCGGGAGTTCAAGGCTTTTGCCCATGCTCAATGGAAAGCACAAGTAC 300

Qy    326 gaagtatatattctctattaatgggttcatacaatgacatttttgacttatttttttcaa 385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GAAGTATATATTCTCTATTAATGGGTTCATACAATGACATTTTTGACTTATTTTTTTCAA 360

Qy    386 attgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagagttcttttg 445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATTGCAGGAGTTTGTTTAAAGACAAAAAATTAGTAGAATACTTAAAGGAGAGTTCTTTTG 420

Qy    446 atgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaatatttctccc 505
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ATGCAGTGTTTCTCGATCCTTTTGATAACTGTGGCTTAATTGTTGCCAAATATTTCTCCC 480

Qy    506 tcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaaggtgcacagt 565
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    481 TCCCCTCCGTGGTCTTCGCCA-GGGAATACTTTGCCACTATCTTGAAGAAGGTGCACAGT 539

Qy    566 gccctgctcctctttcctatgtccccagaattctcttagggttctcagatgccatgac-- 623
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 GCCCTGCTCCTCTTTCCTATGTCCCCAGAATTCTCTTAGGGTTCTCAGATGCCATGACCT 599

Qy    624 tttcaaggagagagtacggaaccacatcatgcacttgg-aggaacatttattatgccacc 682
          ||  | |   |||||||||||| |||||||||||||||||||||||||||||||||||
Db    600 TTCAAGGAGCAGAGTACGGAACCTCATCATGCACTTGGCAGGAACATTTATTATGCCACC 659
```

```
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM10407  row: m  column: 12
           High quality sequence stop: 746.
FEATURES             Location/Qualifiers
     source          1. .882
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4516403"
                     /clone_lib="NIH_MGC_93"
                     /tissue_type="transitional cell papilloma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      259 a    155 c    202 g    266 t
ORIGIN


  Query Match             31.1%;  Score 722.4;  DB 10;  Length 882;
  Best Local Similarity   94.5%;  Pred. No. 3.3e-178;
  Matches  792;  Conservative    0;  Mismatches   41;  Indels    5;  Gaps    4;

Qy   1447 gacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcgtg 1506
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGTG 60

Qy   1507 ctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttgggg 1566
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGG 120

Qy   1567 aaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtgggaaa 1626
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGAAA 180

Qy   1627 taaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgtt 1686
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTGTT 240

Qy   1687 aaattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtgct 1746
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTGCT 300

Qy   1747 ttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaatat 1806
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAATAT 360

Qy   1807 gtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggacgt 1866
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGACGT 420

Qy   1867 gcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttgggagt 1926
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGAGT 480

Qy   1927 gcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatcttatc 1986
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTATC 540

Qy   1987 ttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatctc 2046
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCTCATCTC 600

Qy   2047 tgtcctgcttcataggtgccaccttgtgt-gtttaaagaagggaagctttgtacctttag 2105
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    601 TGTCCTGCTTCATAGGTGCCACCTTGTGTGGTTTAAAGAAGGGAAGCTTTGTACCTTTAG 660

Qy   2106 agtgtaggtg-aaatgaatgaatggcttggagtgcactgagaacagcatatgatttct-t 2163
           |||||||||| |||||||||||||| | |  ||||||||||||||||||||||||||| |
Db    661 AGTGTAGGTGAAAATGAATGAATGGTTGGAGGTGCACTGAGAACAGCATATGATTTCTGG 720

Qy   2164 gctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatc 2223
           |||||||||||||||||||||||||||| || | | |||  |     |||   | |  |||
Db    721 GCTTTGGGGAAAAAGAATGATGCTATGGAAATTGGGGGGTGGTGGATTGGAGAAAATATC 780

Qy   2224 attgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtg 2281
           ||   ||||||||||| |  |    |||||||||||||| || | |    |   | |||
Db    781 ATGCTTTATGTCAAAGGAGG--TGATTTGATAAAAACCAAATACAGGTATTAAAGGTG 836


RESULT   6
BI759297
LOCUS       BI759297                876 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603042939F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5183433 5',
            mRNA sequence.
ACCESSION   BI759297
VERSION     BI759297.1  GI:15750875
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 876)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11457  row: n  column: 10
            High quality sequence stop: 735.
FEATURES             Location/Qualifiers
     source          1. .876
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5183433"
                     /clone_lib="NIH_MGC_116"
                     /lab_host="DH10B"
                     /note="Organ: pooled colon, kidney, stomach; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 3 colons, age 26 yo male, 49 yo
                     female, 71 yo male colon; 46 yo male kidney, and pool of 2
                     stomachs, 62 yo male and 70 yo female.  Library is
                     oligo-dT primed and  directionally cloned (EcoRV site is
                     destroyed upon cloning).  Average insert size 1.4 kb,
                     insert size range 1-3 kb.  Library is normalized and
                     enriched for full-length clones and was constructed by C.
                     Gruber  (Invitrogen).  Research Genetics tracking code
                     023. Note: this is a NIH_MGC Library."
BASE COUNT      226 a    193 c    205 g    252 t
ORIGIN


  Query Match             31.1%;  Score 720.8;  DB 10;  Length 876;
  Best Local Similarity   94.6%;  Pred. No. 8.6e-178;
  Matches  823;  Conservative    0;  Mismatches   37;  Indels   10;  Gaps    7;

Qy      1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7 CTCAGCTGCAGTTCTCTGATGGCTTGCACAGGGTGGACCAGCCCCCTTCCTCTATGTGTG 66

Qy     61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     67 TGTCTGCTGCTGACCTGTGGCTTTGCCGAGGCAGGCAAGCTACTGGTAGTGCCCATGGAT 126

Qy    121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    127 GGGAGCCACTGGTTCACCATGAGGTCGGTGGTGGAGAAACTCATTCTCAGGGGCATGAG 186
```

```
Qy   1676 gaatcagtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagc 1735
          ||  |||||||||| ||| || || || ||| ||||||| ||  |   |  | ||
Db   1689 -AACAAGTGTTAAATCCATATTGTTTTTGTTAGGGAAATAATTCACCATACATTATACAT 1747

Qy   1736 cccagagtgctttaaaaaattctctta--aataaaaataatagactcgctagtcagtaaa 1793
          |  |    |  ||| |||  |  ||  |||  || |  | | | |  ||||  ||
Db   1748 TCAGCACATTTAAAAATAATAATAATAACAATCTAATTGCTGGCCACACCCATCAGGGAA 1807

Qy   1794 gatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatt 1853
          | |  | | ||||||   ||| |   ||| || ||||   ||  ||  ||   ||||||
Db   1808 GGTTCTAGTATATGTGATGTG-CTTTTCCAGTATCTTCAGTCTAGACAACTCTTGCCATC 1866

Qy   1854 tttcagaggacgtgcagacaggctggcattctagatt 1890
          | |  |   | | |||| || |||||| |||   ||
Db   1867 TGT-TGGTAATTTACAGAAAGTCTGGCACTCTGCTTT 1902


RESULT   4
BG402908
LOCUS       BG402908                868 bp    mRNA    linear   EST 12-MAR-2001
DEFINITION  602418710F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4525755 5',
            mRNA sequence.
ACCESSION   BG402908
VERSION     BG402908.1  GI:13296356
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 868)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10432  row: c  column: 04
            High quality sequence stop: 785.
FEATURES             Location/Qualifiers
     source          1. .868
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4525755"
                     /clone_lib="NIH_MGC_93"
                     /tissue_type="transitional cell papilloma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      258 a    151 c    187 g    272 t
ORIGIN


  Query Match             31.9%;  Score 741.2;  DB 10;  Length 868;
  Best Local Similarity   95.9%;  Pred. No. 3.8e-183;
  Matches  804;  Conservative    0;  Mismatches   28;  Indels    6;  Gaps    4;

Qy   1487 gtttcctcttggccgtcgtgctgacagtggccttcatcaccttaaatgttgtgcttatg 1546
          |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTTTCCTCTGGGCCGTCGTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATG 60

Qy   1547 gctaccggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaagaccc 1606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCTACCGGAAATGCTTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCC 120

Qy   1607 attgagaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttccaaac 1666
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTGAGAAGTGGGTGGGAAATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAAC 180

Qy   1667 ttgaaaacagaatcagtgttaaattcattttattcttattaaggaaatactttgcataaa 1726
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTGAAAACAGAATCAGTGTTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAA 240

Qy   1727 ttaatcagccccagagtgctttaaaaaattctcttaaataaaaataatagactcgctagt 1786
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTAATCAGCCCCAGAGTGCTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGT 300

Qy   1787 cagtaaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgacatg 1846
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CAGTAAAGATATTTGAATATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATG 360

Qy   1847 tgccatttttcagaggacgtgcagacaggctggcattctagattacttttcttactctga 1906
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGCCATTTTCAGAGGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGA 420

Qy   1907 aacatggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaa 1966
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AACATGGCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAA 480

Qy   1967 atggcagtttttaatcttatcttttggcttctgcagatggttgcaattgatccttaaccaa 2026
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAA 540

Qy   2027 taatggtcagtcctcatctctgtcctgcttcataggtgccaccttgtgtgtttaaagaag 2086
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TAATGGTCAGTCCTCATCTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAG 600

Qy   2087 ggaagctttgtacctttagagtgtaggtgaaatgaatgaatggcttggagtgcactgaga 2146
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATGGTTTGGAGTGCACTGAGA 660

Qy   2147 acag-catatga-tttcttgctttggggaaaaagaatgatgctatgaaattggtgggtgg 2204
          |||| ||||||| |||||||||||||||||||||||||||||||| ||||||||||
Db    661 ACAGACATATGATTTTCTTGCTTTGGGGAAAAAGAATGATGCTATGAAAATGGTGGGTGG 720

Qy   2205 tgta--tttgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaaaccc 2262
          ||||  ||    |||||||||  |  ||||||| ||  | | |||||||||||||||||
Db    721 TGTAATTTGAGAAAGATAATCCATTGTTATGTCCAAATGGGGTGAATTTGATAAAAACCC 780

Qy   2263 aaaatacagctatgaa--gtgctgggcaagtttacttttttttctgatgtttcctacaa 2318
          ||||      ||||||      |||||||||| |||||||| ||||||||||||||||
Db    781 AAAAATACAGTATGAAAGTGCTTGGGCAAGTTAACTTTTTTCCTGATGTTTCCTACAA 838


RESULT   5
BG291462
LOCUS       BG291462                882 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602387347F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4516403 5',
            mRNA sequence.
ACCESSION   BG291462
VERSION     BG291462.1  GI:13049372
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 882)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
```

```
                     /db_xref="MGD:MGI:1910585"
                     /db_xref="taxon:10090"
                     /clone="0610013P07"
                     /sex="male"
                     /tissue_type="kidney"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     gene            28. .1629
                     /gene="Ugt1a1"
     CDS             28. .1629
                     /gene="Ugt1a1"
                     /note="UDP-glucuronosyltransferase 1 family, member 1
                     data source:MGD, source key:MGI:98898, evidence:ISS
                     putative"
                     /codon_start=1
                     /protein_id="BAB22243.1"
                     /db_xref="GI:12832754"
                     /translation="MDTGLCVPLRGISGLLLLLCALPWAEGAKVLVLPMEGSQWLSMR
                     DVVRELHARGHQTVVLASEVTVHIKGEDFFTLKTYAFPYTKEEYQQEILSDIEKTFKT
                     QHFVKAFFETTASIRNFFDLYSNSCIALLHNKMLIQQLNSSFFDVILTDPIFPCGAVL
                     AKYLQIPAVFILRSLSCGIEYEATQCPNPSSYIPNLLTRLSDHMDFLQRVQNMLYYLV
                     LKYICRLSITPYESLASELLQREVSLVEVLSHASVWLFRGDFVLDYPRPIMPNMVFIG
                     GINCVTKKPLSQEFEAYVNASGEHGIVVFSLGSMVSEIPEKKAMEIAEALGRIPQTVL
                     WRYTGTRPSNLAKNTILVKWLPQNDLLGHPKTRAFITHSGSHGIYEGICNGVPMVMMP
                     LFGDQMDNAKRMETRGAGVTLNVLEMTADDLENALKTVINNKSYKENIMRLSSLHKDR
                     PIEPLDLAVFWVEYVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAIVLTVVFIVFK
                     CCAYGCRKCFGGKGRVKKSHKSKTH"
     polyA_signal    2183. .2188
                     /note="putative"
     polyA_site      2202
                     /note="putative"
BASE COUNT      568 a    514 c    514 g    606 t
ORIGIN
```

```
  Query Match             33.6%;  Score 780.2;  DB 11;  Length 2202;
  Best Local Similarity   65.9%;  Pred. No. 3.1e-193;
  Matches 1211;  Conservative    0;  Mismatches  613;  Indels   13;  Gaps    5;

Qy      59 tgtgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatgg 118
           ||   ||||||    ||     ||  || || |  |  ||  | |||||  | |||||||
Db      74 TGCTCCTGCTGTGTGCCCTGCCCTGGGCTGAAGGTGCAAAAGTGCTGGTGTTACCCATGG 133

Qy     119 atgggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggggcatg 178
           | || ||||| ||| | | |||||||    || |||  | | |||     | | || ||
Db     134 AGGGCAGCCAGTGGCTGAGCATGAGGGATGTTGTGCGGGAGCTCCACGCCCGAGGTCACC 193

Qy     179 aggtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacag 238
           ||   || ||  |     |||||||||  |  |||  |   | ||  |       | |||
Db     194 AGACTGTGGTCCTGGCTTCAGAGGTGACTGTGCACATCAAAGGAGAAGACTTCTTCACCC 253

Qy     239 tgaagacttattcaacttcatataccctggagg---atctggaccgggagttcaaggctt 295
           | || || ||| |    | ||||||||  ||| |   ||| | |         | |  |
Db     254 TCAAAACCTATGCCTTTCCATATACCAAGGAAGAATATCAGCAAGAAATACTGAGCGACA 313

Qy     296 ttgcccatgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcat 355
           |||   |  |       | |  |||  |    |      | || |  |   |  |  ||
Db     314 TTGAGAAGACCTTTAAAACACAACATTTTGTGAAGGCCTTTTTTGAAACTACAGCATCTA 373

Qy     356 acaatgacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaat 415
             |   || ||||||||  |  |  |       ||| |    ||| ||  |  | || |
Db     374 TAAGAAACTTTTTTGACCTCTACTCAAATTCTTGTATAGCTTTATTGCACAATAAGATGC 433

Qy     416 tagtagaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataact 475
           |  |  |  |  | ||    |||| ||||||||   | ||    || ||  ||      ||
Db     434 TGATCCAGCAACTGAATTCCAGTTTTTTTGATGTGATCTTAACAGACCCCATTTTCCCCT 493

Qy     476 gtggcttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatac 535
           ||||   | |  | ||||| ||  |      | ||  | ||| |  |     |     |
Db     494 GTGGAGCAGTACTGGCCAAGTACCTACAGATTCCTGCTGTGTTTATTCTGCGTTCTCTTT 553

Qy     536 tttgccactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaa 595
             ||   |       | || |   |||| || ||   ||||   || ||| | || |
```

```
Db     554 CCTGTGGCATAGAATATGAGGCCACACAATGTCCAAATCCTTCCTCTTATATTCCGAACC 613

Qy     596 ttctcttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgc 655
           | |||  | ||  | ||  ||   ||||   |||  | | ||  || |  ||||    |
Db     614 TACTCACAAGGCTTTCTGACCACATGGACTTCCTGCAAAGGGTCCAGAACATGCTGTACT 673

Qy     656 acttggaggaacatttattatgccaccgtttttttcaaaaatgccctagaaatagcctctg 715
           |  |||      | |   | ||||     |   |||          |    | |||||||
Db     674 ATCTGGTCCTGAAGTACATTTGCCGTTTATCAATCACTCCCTATGAAAGCCTGGCCTCTG 733

Qy     716 aaattctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgt 775
           |  || | || | |   ||  |       ||  |   ||||||  ||||  | ||| |||
Db     734 AGCTTTTGCAGAGAGAAGTATCCTTAGTGGAGGTTCTCAGCCATGCATCCGTGTGGCTGT 793

Qy     776 tgcgaacggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattg 835
           | |||  |||||||||  | |||||  ||||  |||  | |||||  ||||||  |||||||||
Db     794 TCCGAGGGGACTTTGTACTCGACTACCCCAGGCCCATCATGCCTAACATGGTCTTCATTG 853

Qy     836 gtggtatcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatg 895
           | || || |||||   |      |||||  |  ||  ||||||||||||||||  | || |
Db     854 GGGGCATAAACTGTGTTACCAAGAAGCCCCTCTCTCAGGAATTTGAAGCCTATGTCAACG 913

Qy     896 cttctggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccag 955
           | ||||| || ||||| || || ||||||||||||||||||  ||||||||||| ||||| |
Db     914 CCTCTGGGGAGCATGGCATCGTTGTTTTCTCTTTGGGATCCATGGTCTCAGAGATTCCGG 973

Qy     956 agaagaaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggc 1015
           ||||||||| |||| |||||||||| |||||||||| ||| |||||||| ||||||||||
Db     974 AGAAGAAAGCCATGGAAATTGCTGAGGCTTTGGGCAGAATTCCTCAGACGGTCCTGTGGC 1033

Qy    1016 ggtacactggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctac 1075
           | ||||| |||||  |||||||||||||||| || ||||| || ||||| || |||||||
Db    1034 GCTACACCGGAACTAGACCATCGAATCTTGCAAAGAACACAATTCTTGTCAAATGGCTAC 1093

Qy    1076 cccaaaacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttccc 1135
           ||||||| |||||||||||||| || | ||| || || || ||||| ||  |||| ||||
Db    1094 CCCAAAATGATCTGCTTGGTCATCCAAAGACTCGGGCATTCATCACACACTCTGGCTCCC 1153

Qy    1136 atggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtg 1195
           ||||| |||||||| | |||||||||| ||||| ||||||||||||||| | ||||| |
Db    1154 ATGGTATTTATGAAGGAATATGCAATGGAGTTCCGATGGTGATGATGCCCCTATTTGGCG 1213

Qy    1196 atcagatggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttc 1255
           |||||||||||||||| |||||||||| |||  |||||||| ||||||||||||||| |
Db    1214 ATCAGATGGACAATGCCAAGCGCATGGAAACTCGGGGAGCTGGGGTGACCCTGAATGTCC 1273

Qy    1256 tggaaatgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagtt 1315
           | |||||||| |||| ||||| |||||||| || ||| | ||||||||  |||| || |
Db    1274 TTGAAATGACTGCTGATGATTTGGAAAATGCCCTTAAAACTGTCATCAACAACAAGAGCT 1333

Qy    1316 acaaggagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctgg 1375
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||  | ||||| ||||
Db    1334 ACAAGGAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGTCCTATAGAGCCTCTGG 1393

Qy    1376 acctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgcc 1435
           ||||||| |||||||||||||| |  |||||||||||||||| || ||||||||||||
Db    1394 ACCTGGCTGTGTTCTGGGTGGAATACGTGATGAGGCACAAGGGGGCACCACACCTGCGCC 1453

Qy    1436 ccgcagcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctct 1495
           | || ||||| |||||||||||||| |||||||| |||||||| |||||||| ||||||
Db    1454 CGGCCGCCCATGACCTCACCTGGTATCAGTACCACTCCTTGGATGTGATTGGCTTCCTCC 1513

Qy    1496 tggccgtcgtgctgacagtggccttcatcaccttaaatgttgtgcttatggctaccgga 1555
           ||||| | ||| |||||||| ||||||   |||||||||||||| ||||||| |||||
Db    1514 TGGCCATTGTGTTGACAGTGGTCTTCATTGTCTTTAAATGTTGTGCCTATGGCTGCCGGA 1573

Qy    1556 aatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaag 1615
           ||||||| |||  ||| |||||||| |||||| | ||||||||||||||||||||||||
Db    1574 AATGCTTTGGGGGAAAGGGGCGAGTGAAGAAATCACACAAATCCAAGACCCATTGAGAAG 1633

Qy    1616 tgggtgggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaaca 1675
           |||| || |   ||||  || | ||    |||| ||  ||| ||  ||  | ||
Db    1634 TGGGGGGAAGTGAAGGAGAAGTATTAGTTCATTATCTGATCAGTTGAAACTTGGA----- 1688
```

```
Qy   1073 taccccaaaacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggtt 1132
          |||||||||| |||||||||||||| || | ||| || || || ||||| ||  |||| |
Db   1180 TACCCCAAAATGATCTGCTTGGTCATCCAAAGACTCGGGCATTCATCACACACTCTGGCT 1239

Qy   1133 cccatggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttg 1192
          |||||||| |||||||| | |||||||||||| ||||| |||||||||||||||| | ||||
Db   1240 CCCATGGTATTTATGAAGGAATATGCAATGGAGTTCCGATGGTGATGATGCCCCTATTTG 1299

Qy   1193 gtgatcagatggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatg 1252
          | |||||||||||||||| |||||||||| ||| |||||||| ||||||||||||||
Db   1300 GCGATCAGATGGACAATGCCAAGCGCATGGAAACTCGGGGAGCTGGGGTGACCCTGAATG 1359

Qy   1253 ttctggaaatgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaa 1312
          | || ||||||||| |||| ||||| |||||||| || ||| | ||||||||  |||| |
Db   1360 TCCTTGAAATGACTGCTGATGATTTGGAAAATGCCCTTAAAACTGTCATCAACAACAAGA 1419

Qy   1313 gttacaaggagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgc 1372
          | |||||||||||||||||||||||||||||||||||||||||||||||| ||  | ||||| |
Db   1420 GCTACAAGGAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGTCCTATAGAGCCTC 1479

Qy   1373 tggacctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgc 1432
          |||||||||| |||||||||||||| |  |||||||||||||||| || ||||||||||
Db   1480 TGGACCTGGCTGTGTTCTGGGTGGAATACGTGATGAGGCACAAGGGGGCACCACACCTGC 1539

Qy   1433 gccccgcagcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcc 1492
          |||| || ||||| |||||||||||||| |||||||| |||||||| |||||||| ||||
Db   1540 GCCCGGCCGCCCATGACCTCACCTGGTATCAGTACCACTCCTTGGATGTGATTGGCTTCC 1599

Qy   1493 tcttggccgtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctacc 1552
          || ||||| | ||| |||||||| ||||||   ||||||||||||||| ||||||| ||
Db   1600 TCCTGGCCATTGTGTTGACAGTGGTCTTCATTGTCTTTAAATGTTGTGCCTATGGCTGCC 1659

Qy   1553 ggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgag 1612
          |||||||||| |||  ||| |||||||| |||||| | ||||||||||||||||||||||
Db   1660 GGAAATGCTTTGGGGGAAAGGGGCGAGTGAAGAAATCACACAAATCCAAGACCCATTGAG 1719

Qy   1613 aagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaa 1672
          ||||||| || |   ||||  || | ||    ||||  ||  ||| ||  ||  | ||
Db   1720 AAGTGGGGGGAAGTGAAGGAGAAGTATTAGTTCATTATCTGATCAGTTGAAACTTGGA-- 1777

Qy   1673 acagaatcagtgttaaattcattttattcttattaaggaaatactttgcataaattaatc 1732
              ||  ||||||| || ||| || || || ||| ||||||| ||  |   |  | ||
Db   1778 ----AACAAGTGTTAGATCCATATTGTTTTTGTTAGGGAAATAATTCACCATACATTATA 1833

Qy   1733 agccccagagtgctttaaaaaattctctta--aataaaaataatagactcgctagtcagt 1790
               |   |  | ||| |||  |  ||  |||  || |  | | | | |  ||||
Db   1834 CATTCAGCACATTTAAAAATAATAATAATAACAATCTAATTGCTGGCCACACCCATCAGG 1893

Qy   1791 aaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgcc 1850
           ||| |  | | ||||||   ||| |   ||| || ||||   ||  ||  ||   ||||
Db   1894 GAAGGTTCTAGTATATGTGATGTG-CTTTTCCAGTATCTTCAGTCTAGACAACTCTTGCC 1952

Qy   1851 atttttcagaggacgtgcagacaggctggcattctagatt 1890
          || | |  |   |  | |||| || ||||||| |||   ||
Db   1953 ATCTGT-TGGTAATTTACAGAAAGTCTGGCACTCTGCTTT 1991
```

RESULT 3
AK002629

```
LOCUS       AK002629               2202 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus adult male kidney cDNA, RIKEN full-length enriched
            library, clone:0610013P07:UDP-glucuronosyltransferase 1 family,
            member 1, full insert sequence.
ACCESSION   AK002629
VERSION     AK002629.1  GI:12832753
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:0610013P07.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 2202)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0. Second strand cDNA
            was prepared with the primer adapter of sequence[5'
            GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was
            cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
            XhoI. Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .2202
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
```

```
Db   2048 AAATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACC 2107

Qy   2025 aataatggtcagtcctcatctctgtcctgcttcataggtgccaccttgtgtgtttaaaga 2084
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2108 AATAATGGTCAGTCCTCATCTCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGA 2167

Qy   2085 agggaagctttgtacctttagagtgtaggtgaaatgaatgaatggcttggagtgcactga 2144
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2168 AGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGGCTTGGAGTGCACTGA 2227

Qy   2145 gaacagcatatgatttcttgctttggggaaaaagaatgatgctatgaaattggtgggtgg 2204
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2228 GAACAGCATATGATTTCTTGCTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGG 2287

Qy   2205 tgtatttgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaacccaa 2264
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2288 TGTATTTGAGAAGATAATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAA 2347

Qy   2265 aatacagctatgaagtgctgggcaagtttacttttttctgatgtttcctacaact 2320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2348 AATACAGCTATGAAGTGCTGGGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 2403


RESULT   2
BC012716
LOCUS       BC012716               2306 bp    mRNA    linear   HTC 20-AUG-2001
DEFINITION  Mus musculus, Similar to UDP-glucuronosyltransferase 1 family,
            member 1, clone IMAGE:3993384, mRNA.
ACCESSION   BC012716
VERSION     BC012716.1  GI:15215247
KEYWORDS    HTC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 18 Row: e Column: 4
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 801898
            This clone has the following problem: incomplete processing.
FEATURES             Location/Qualifiers
     source          1. .2306
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3993384"
                     /tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
                     old, gross tissue."
                     /clone_lib="NCI_CGAP_Mam5"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
BASE COUNT      594 a    559 c    528 g    625 t
ORIGIN
```

```
  Query Match           34.0%;  Score 789.6;  DB 11;  Length 2306;
  Best Local Similarity  66.6%;  Pred. No. 1.1e-195;
  Matches 1225;  Conservative    0;  Mismatches  599;  Indels   16;  Gaps    6;

Qy     56 gtgtgtgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgccca 115
          | || | ||| |      | || | ||  |    |  | ||||| ||||| || |||
Db    163 GGGTTTCTCTTCCTAGTGCTTGGGCCTCAGTTCTAGGTGACAAGCTGCTGGTGGTCCCCC 222

Qy    116 tggatgggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggc 175
           |||||| |||||||||  | | ||||| |  | | |||||  | |||| |    | || |
Db    223 AGGATGGAAGCCACTGGCTTAGCATGAAGGAGATAGTGGAACATCTCAGTGAACGAGGAC 282

Qy    176 atgaggtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgca 235
          ||||  |  | ||  |  |||||||  || | ||  ||   |||||  |||     | | ||
Db    283 ATGACATTATGGTCCTAGTGCCAGAAGTCAATTTGCTTTTGGGAGAATCCAAATACTACA 342

Qy    236 cagtgaagacttattcaacttcatataccctggaggatctggaccgggagttcaaggctt 295
              ||| |  |  |||  | | || | ||| |||||  ||  |        ||      | |
Db    343 GGAGGAAAATCTTCTCAGTTACCTACAGCCTAGAGGAACTACAGACCCGTTTTCGCACCT 402

Qy    296 ttgcccatgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcat 355
          |||   |    || |       |    |         ||   | |||||  |  |
Db    403 TTGGAAACAACCACTTTCTTCCCAGTGCCGCCCTGATTGGTCCTCTAAGAGAGTACAGGA 462

Qy    356 acaatgacatt---tttgacttatttttttcaaattgcaggagtttgtttaaagacaaaa 412
          ||||    |||    | ||| | |  ||||  |  |||  |||  |  | | || |||
Db    463 ACAACATGATTGTCGTGGACATGTGCTTTTTCAGCTGCCAGAGCCTCCTGAAGGACTCTG 522

Qy    413 aattagtagaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgata 472
                |    | | | |||||     ||||||||  ||||    || ||
Db    523 CCACCCTCAGCTTCCTCAGGGAGAACAAGTTTGATGCTCTGTTCACAGACCCAGCCATGC 582

Qy    473 actgtggcttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaa 532
           ||||||  | ||  | ||  | ||| ||  ||| |||||  ||   | ||  ||| ||
Db    583 CCTGTGGTGTGATCCTGGCTGAGTATCTCAACCTGCCCTCTGTCTACCTCTTCAGAGGTT 642

Qy    533 tactttgccactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtcccca 592
          | |  |||           |      ||  ||  ||||    ||| | ||||| || ||||
Db    643 TCCCATGCTCTCTGGAACACATGCTTGGTCAAAGCCCAAGCCCTGTGTCCTACGTTCCCA 702

Qy    593 gaattctcttagggttctcagatgccatgactttcaaggagagagtacggaaccacatca 652
          || |   |      ||||||||   ||||||  ||      |  |  |    |||   ||
Db    703 GATTCTACACCAAATTCTCAGACCACATGACATTTCCCCAACGGCTGGCCAACTTTATTG 762

Qy    653 tgcacttggaggaacatttattatgccaccgtttttttcaaaaatgccctagaaatagcct 712
          |  || |   |||| | |   |||  |  || | |    |||      |   || ||||
Db    763 TTAACATCTTGGAAAACTACCTATATTATTGTCTGTATTCAAAGTATGAAATCATTGCCT 822

Qy    713 ctgaaattctccaaacacctgttacggagtatgatctctacagccacacatcaatttggt 772
          | ||  | ||| | | |  |||  |     |   |  |  ||| |  ||  | |||
Db    823 CAGACCTCCTCAAGAGAGATGTGTCCCTACCTTCCTTACA---CCAGAATTCTCTGTGGC 879

Qy    773 tgttgcgaacggactttgttttggactatcccaaacccgtgatgcccaacatgatcttca 832
          |||| ||    || ||||| || || |||||||  || || |||||||||||||||||
Db    880 TGTTACGGTACGATTTTGTGTTCGAATATCCCAGGCCAGTCATGCCCAACATGATCTTCA 939

Qy    833 ttggtggtatcaactgccatcagggaaagccgttgcctatggaatttgaagcctacatta 892
          |||| || ||||||||  |  ||        | || ||  ||||||||||||||  | |
Db    940 TTGGAGGGATCAACTGTAAGAAGAAGGGAAAGCTGACTCAGGAATTTGAAGCCTATGTCA 999

Qy    893 atgcttctggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattc 952
          | || ||||| || ||||| || |||||||||||||||||||| |||||||||| ||||
Db   1000 ACGCCTCTGGGGAGCATGGCATCGTGGTTTTCTCTTTGGGATCCATGGTCTCAGAGATTC 1059

Qy    953 cagagaagaaaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgt 1012
          | |||||||||| |||| |||||||| |||||||| ||| ||||||| |||||||
Db   1060 CGGAGAAGAAAGCCATGGAAATTGCTGAGGCTTTGGGCAGAATTCCTCAGACGGTCCTGT 1119

Qy   1013 ggcggtacactggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggc 1072
          |||| ||||| ||||| ||||||||||||||| || ||||| || ||||| || ||||
Db   1120 GGCGCTACACCGGAACTAGACCATCGAATCTTGCAAAGAACACAATTCTTGTCAAATGGC 1179
```

```
     source          1. .2450
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3854370"
                     /tissue_type="Colon, adenocarcinoma"
                     /clone_lib="NIH_MGC_65"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
BASE COUNT      669 a    509 c    565 g    707 t
ORIGIN
```

Query Match 70.3%; Score 1631.6; DB 11; Length 2450;
Best Local Similarity 82.6%; Pred. No. 0;
Matches 1880; Conservative 0; Mismatches 394; Indels 2; Gaps 1;

```
Qy      47 ttcctctatgtgtgtgtctgctgctgacctgtggctttgccgaggcagggaagctactgg 106
           || ||  | | || |   |  |       || ||| | |  | |  |  |||| ||||
Db     128 TTTCTGCAGGGGTTTTCTTCTTAGCACTTTGGGGCATGGTTGTAGGTGACAAGCTGCTGG 187

Qy     107 tagtgcccatggatgggagccactggttcaccatgaggtcggtggtggagaaactcattc 166
           | || ||   ||| || |||||||| | |  |||| |    | || |||   |||| |
Db     188 TGGTCCCTCAGGACGGAAGCCACTGGCTTAGTATGAAGGATATAGTTGAGGTTCTCAGTG 247

Qy     167 tcaggggcatgaggtggttgtagtcatgccagaggtgagttggcaactgggaagatcac 226
           | |||| ||||||  | || || ||  |||| || || | || ||   ||  |  |||
Db     248 ACCGGGGTCATGAGATTGTAGTGGTGGTGCCTGAAGTTAATTTGCTTTTGAAAGAATCCA 307

Qy     227 tgaattgcacagtgaagacttattcaacttcatataccctggaggatctggaccgggagt 286
           | | ||||   || |  ||| ||    | |||  ||  || || ||| |       |
Db     308 AATACTACACAAGAAAAATCTATCCAGTGCCGTATGACCAAGAAGAGCTGAAGAACCGTT 367

Qy     287 tcaaggcttttgcccatgctcaatgg--aaagcacaagtacgaagtatatattctctatt 344
           | |  | ||||   |   ||| | |   |   |  | | || |     |     |
Db     368 ACCAATCATTTGGAAACAATCACTTGCTGAGCGATCATTCCTAACTGCTCCTCAGACAGA 427

Qy     345 aatgggttcatacaatgacatttttgacttatttttttcaaattgcaggagtttgtttaa 404
           |     |||       || ||| | | |  ||    || |||  |||  |  |  |
Db     428 GTACAGGAATAACATGATTGTTATTGGCCTGTACTTCATCAACTGCCAGAGCCTCCTGCA 487

Qy     405 agacaaaaaattagtagaatacttaaaggagagttcttttgatgcagtgtttctcgatcc 464
           ||||  |     |  | | ||| ||||||||    ||||||||  | ||    || ||
Db     488 GGACAGGGACACCCTGAACTTCTTTAAGGAGAGCAAGTTTGATGCTCTTTTCACAGACCC 547

Qy     465 ttttgataactgtggcttaattgttgccaaatatttctccctcccctccgtggtcttcgc 524
                |||||  | ||  | ||  | ||||   ||| || || |||  | ||
Db     548 AGCCTTACCCTGTGGGGTGATCCTGGCTGAGTATTTGGGCCTACCATCTGTGTACCTCTT 607

Qy     525 caggggaatactttgccactatcttgaagaaggtgcacagtgccctgctcctctttccta 584
           ||||||  | |  ||  |       |  |          |||| |  ||| | |||||
Db     608 CAGGGGTTTTCCGTGTTCCCTGGAGCATACATTCAGCAGAAGCCCAGACCCTGTGTCCTA 667

Qy     585 tgtccccagaattctcttagggttctcagatgccatgactttcaaggagagagtacggaa 644
           | |||||      |  |  ||| |||||   |||||||||     |  ||||    ||
Db     668 CATTCCCAGGTGCTACACAAAGTTTTCAGACCACATGACTTTTTCCCAACGAGTGGCCAA 727

Qy     645 ccacatcatgcacttggaggaacatttattatgccaccgttttttcaaaaatgccctaga 704
           |  | |  |  | |||  ||| |  |   |||   | || | ||   |||      |||
Db     728 CTTCCTTGTTAATTTGTTGGAGCCCTATCTATTTTATTGTCTGTTTTCAAAGTATGAAGA 787

Qy     705 aatagcctctgaaattctccaaacacctgttacggagtatgatctctacagccacacatc 764
           | | || || |   | ||| | | |  |||              | ||      |    ||
Db     788 ACTCGCATCAGCTGTCCTCAAGAGAGATGTGGATATAATCACCTTATATCAGAAGGTCTC 847

Qy     765 aatttggttgttgcgaacggactttgttttggactatcccaaacccgtgatgcccaacat 824
           |||||  ||||  ||  ||||||||  | || |||||  |  || || ||||||||||
Db     848 TGTTTGGCTGTTAAGATATGACTTTGTGCTTGAATATCCTAGGCCGGTCATGCCCAACAT 907

Qy     825 gatcttcattggtggtatcaactgccatcagggaaagccgttgcctatggaatttgaagc 884
           | |||||||||| ||||||||||  |  || | ||    ||| ||  ||||||||||||
Db     908 GGTCTTCATTGGAGGTATCAACTGTAAGAAGAGGAAAGACTTGTCTCAGGAATTTGAAGC 967

Qy     885 ctacattaatgcttctggagaacatggaattgtggttttctctttgggatcaatggtctc 944
```

```
                                                                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     968 CTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTC 1027

Qy     945 agaaattccagagaagaaagctatggcaattgctgatgctttgggcaaaatccctcagac 1004
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1028 AGAAATTCCAGAGAAGAAAGCTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGAC 1087

Qy    1005 agtcctgtggcggtacactggaacccgaccatcgaatcttgcgaacaacacgatacttgt 1064
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1088 AGTCCTGTGGCGGTACACTGGAACCCGACCATCGAATCTTGCGAACAACACGATACTTGT 1147

Qy    1065 taagtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcctttatcaccca 1124
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1148 TAAGTGGCTACCCCAAAACGATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCA 1207

Qy    1125 tgctggttcccatggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgcc 1184
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1208 TGCTGGTTCCCATGGTGTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCC 1267

Qy    1185 cttgtttggtgatcagatggacaatgcaaagcgcatggagactaagggagctggagtgac 1244
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1268 CTTGTTTGGTGATCAGATGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGAC 1327

Qy    1245 cctgaatgttctggaaatgacttctgaagatttagaaaatgctctaaaagcagtcatcaa 1304
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1328 CCTGAATGTTCTGGAAATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAA 1387

Qy    1305 tgacaaaagttacaaggagaacatcatgcgcctctccagccttcacaaggaccgcccggt 1364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1388 TGACAAAAGTTACAAGGAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGT 1447

Qy    1365 ggagccgctggacctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgcc 1424
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1448 GGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCC 1507

Qy    1425 acacctgcgcccgcagcccacgacctcacctggtaccagtaccattccttggacgtgat 1484
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1508 ACACCTGCGCCCCGCAGCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGAT 1567

Qy    1485 tggtttcctcttggccgtcgtgctgacagtggccttcatcacctttaaatgttgtgctta 1544
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1568 TGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTA 1627

Qy    1545 tggctaccggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaagac 1604
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1628 TGGCTACCGGAAATGCTTGGGGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGAC 1687

Qy    1605 ccattgagaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttccaa 1664
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1688 CCATTGAGAAGTGGGTGGGAAATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAA 1747

Qy    1665 acttgaaaaacagaatcagtgttaaattcattttattcttattaaggaaatactttgcata 1724
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1748 ACTTGAAAACAGAATCAGTGTTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATA 1807

Qy    1725 aattaatcagccccagagtgctttaaaaaattctcttaaataaaaataatagactcgcta 1784
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1808 AATTAATCAGCCCCAGAGTGCTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTA 1867

Qy    1785 gtcagtaaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgaca 1844
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1868 GTCAGTAAAGATATTTGAATATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACA 1927

Qy    1845 tgtgccatttttcagaggacgtgcagacaggctggcattctagattacttttcttactct 1904
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1928 TGTGCCATTTTTCAGAGGACGTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCT 1987

Qy    1905 gaaacatggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctactgc 1964
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1988 GAAACATGGCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGC 2047

Qy    1965 aaatggcagtttaatcttatcttttggcttctgcagatggttgcaattgatccttaacc 2024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 21, 2002, 09:38:48 ; Search time 4821.79 Seconds
(without alignments)
6494.048 Million cell updates/sec

Title: US-09-835-082-3
Perfect score: 2320
Sequence: 1 ctcagctgcagttctctgat.....ttctgatgtttcctacaact 2320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1631.6	70.3	2450	11	BC011409	BC011409 Homo sapi
	2	789.6	34.0	2306	11	BC012716	BC012716 Mus muscu
	3	780.2	33.6	2202	11	AK002629	AK002629 Mus muscu
	4	741.2	31.9	868	10	BG402908	BG402908 602418710
	5	722.4	31.1	882	10	BG291462	BG291462 602387347
	6	720.8	31.1	876	10	BI759297	BI759297 603042939
	7	707.4	30.5	884	10	BI762347	BI762347 603049004
c	8	695.8	30.0	818	9	AI307269	AI307269 tb24d12.x
	9	695.4	30.0	881	10	BG289777	BG289777 602385071
	10	680.6	29.3	1005	10	BE870911	BE870911 601448862
	11	679.2	29.3	697	10	BG747143	BG747143 602704364
c	12	669.4	28.9	764	9	AI831562	AI831562 wj39b01.x
	13	644	27.8	723	10	BG424452	BG424452 602447772
	14	641.2	27.6	829	10	BG469506	BG469506 602532818
	15	640	27.6	833	10	BG400591	BG400591 602464244
c	16	639.8	27.6	791	9	AI765716	AI765716 wi83f06.x
	17	636.8	27.4	658	9	AV652040	AV652040 AV652040
	18	635.4	27.4	945	10	BG167885	BG167885 602339857
	19	625.6	27.0	724	10	BF031444	BF031444 601559244
	20	621.8	26.8	688	10	BG286763	BG286763 602382611
	21	618.8	26.7	796	10	BG432363	BG432363 602496688
c	22	604.8	26.1	861	9	AI829268	AI829268 wk58a10.x
	23	601	25.9	647	9	AV692287	AV692287 AV692287
	24	596	25.7	763	10	BG399019	BG399019 602440878
	25	594.2	25.6	663	9	AV646537	AV646537 AV646537
	26	592.4	25.5	859	10	BE249993	BE249993 600943171
c	27	589.8	25.4	640	9	AI635995	AI635995 wa14e12.x
c	28	588	25.3	683	9	AI480360	AI480360 tm51d11.x
	29	584.8	25.2	812	10	BG401258	BG401258 602465359
c	30	584	25.2	621	9	AW151709	AW151709 xf68b10.x
c	31	583	25.1	667	9	AW051715	AW051715 wx27g08.x
	32	583	25.1	969	10	BG291839	BG291839 602386068
c	33	572	24.7	636	9	AI767470	AI767470 wh27a05.x
	34	567.4	24.5	581	10	BE379923	BE379923 601159412
	35	567	24.4	623	10	BG171502	BG171502 602322042
	36	557.8	24.0	743	10	BG386668	BG386668 602454928
	37	556.4	24.0	741	10	BG433491	BG433491 602497921
c	38	550.6	23.7	619	9	AW300312	AW300312 xs59f05.x
	39	549.8	23.7	851	10	BG401340	BG401340 602465455
	40	543.8	23.4	700	10	BF130433	BF130433 601818721
c	41	542.8	23.4	608	9	AW469219	AW469219 hc79e05.x
	42	542.8	23.4	953	10	BG328092	BG328092 602427143
	43	537	23.1	935	10	BF234991	BF234991 602027544
c	44	535.4	23.1	626	9	AW082933	AW082933 xc04f05.x
	45	532.2	22.9	666	10	BF027719	BF027719 601764020

ALIGNMENTS

RESULT 1
BC011409
LOCUS BC011409 2450 bp mRNA linear HTC 30-JUL-2001
DEFINITION Homo sapiens, clone IMAGE:3854370, mRNA.
ACCESSION BC011409
VERSION BC011409.1 GI:15030273
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2450)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalon@bcm.tmc.edu.
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W., Muzny,D.M., Gibbs,R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 13 Row: b Column: 4
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4507814
This clone has the following problem: incomplete processing.
FEATURES Location/Qualifiers

```
;    APPLICANT:  BELKNAP, WILLIAM R
;    TITLE OF INVENTION:  DNA SEQUENCES ENCODING SOLANIDINE
;    TITLE OF INVENTION:  UDP-GLUCOSE GLUCOSYLTRANSFERASE AND USE TO REDUCE
;    TITLE OF INVENTION:  GLYCOALKALOIDS IN SOLANACEOUS PLANTS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NANCY J. PARSONS
;      STREET:  800 BUCHANAN ST.
;      CITY:  ALBANY
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94710
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/797,226
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PARSONS, NANCY J
;      REGISTRATION NUMBER:  40,364
;      REFERENCE/DOCKET NUMBER:  0011.97
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 559-5731
;      TELEFAX:  (510) 559-5777
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1607 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Solanum tuberosum
;      STRAIN:  cv. Lemhi Russet
;      INDIVIDUAL ISOLATE:  SGT1750
;      DEVELOPMENTAL STAGE:  mature, somatic
;      TISSUE TYPE:  tuber
;    IMMEDIATE SOURCE:
;      LIBRARY:  lambda gt11 cDNA library
;      CLONE:  SGT 1750
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20..1486
;      OTHER INFORMATION:  /product= "solanidine
;      OTHER INFORMATION:  glucosyltransferase"
;    PUBLICATION INFORMATION:
;      AUTHORS:  Moehs, Charles P.
;      AUTHORS:  Allen, Paul V.
;      AUTHORS:  Friedman, Mendel
;      AUTHORS:  Belknap, William R.
;      TITLE:  Cloning and expression of solanidine
;      TITLE:  UDP-glucose glucosyltransferase from potato
;      JOURNAL:  The Plant Journal
;      VOLUME:  11
;      ISSUE:  2
;      DATE:  1997
;      RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 488
US-08-797-226-1


  Query Match             1.9%;  Score 44.8;  DB 2;  Length 1607;
  Best Local Similarity   53.4%;  Pred. No. 0.0072;
  Matches   94;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy   1068 gtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcctttatcacccatgc 1127
          |||| | || ||        |  | |  ||  |    ||  | |  || || || |||
Db   1069 GTGGCTCCCACAGCTTACGATCATGGAACATTCAGCAACAGGCGGGTTCATGACTCATTG 1128

Qy   1128 tggttcccatggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgccctt 1187
          |||| |  ||   |||   |||  |||      |||||| || ||| | |    |||  |
Db   1129 TGGTACTAATTCGGTTCTGGAAGCCATCACTTTTGGCGTGCCAATGATAACATGGCCACT 1188

Qy   1188 gtttggtgatcagatggacaatgcaaagcgcatggagactaagggagctggagtga 1243
           | || ||||||  |  |||| |  |||    | |||  || ||||   ||| | |
Db   1189 TTATGCTGATCAATTCTACAACGAGAAGGTAGTCGAGGTTAGGGGATTGGGAATCA 1244


RESULT  15
PCT-US92-00282-37
; Sequence 37, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714527 CUSH
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  51 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-37


  Query Match             1.9%;  Score 43.6;  DB 5;  Length 51;
  Best Local Similarity   92.0%;  Pred. No. 0.0021;
  Matches   46;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy   1444 cacgacctcacctggtaccagtaccattccttggacgtgattggtttcct 1493
           |||||||||||||||| ||||||||| || ||||| ||||||||||||||
Db      1 CACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGTTTCCT 50


Search completed:  August 21, 2002, 11:59:49
Job time:  32837 sec
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  513 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-20


  Query Match             5.8%;  Score 135;  DB 5;  Length 513;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-29;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    873 ggaatttgaagcctacattaatgcttctggagaacatggaattgtggttttctctttggg 932
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    230 GGAATTTGAAGCCTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTCTCTTTGGG 289

Qy    933 atcaatggtctcagaaattccagagaagaaagctatggcaattgctgatgctttgggcaa 992
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    290 ATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATGGCAATTGCTGATGCTTTGGGCAA 349

Qy    993 aatccctcagacagt 1007
          |||||||||||||||
Db    350 AATCCCTCAGACAGT 364


RESULT  13
US-08-522-421-1
; Sequence 1, Application US/08522421
; Patent No. 5908973
;  GENERAL INFORMATION:
;    APPLICANT:  Abu-Bakar, Umi Kalsom
;    APPLICANT:  Barton, Sarah Louise
;    APPLICANT:  Gallego-Veigas, Pedro Pablo
;    APPLICANT:  Gray, Julie Elizabeth
;    APPLICANT:  Grierson, Donald
;    APPLICANT:  Lowe, Alexandra Louise
;    APPLICANT:  Picton, Steve
;    APPLICANT:  Whotton, Lee Colin
;    TITLE OF INVENTION:  DNA, DNA CONSTRUCTS, CELLS AND PLANTS
;    TITLE OF INVENTION:  DERIVED THEREFROM
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pillsbury Madison & Sutro, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/522,421
;      FILING DATE:  11-JAN-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305868.3
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305869.1
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305859.2
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305865.9
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305866.7
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305867.5
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305860.0
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305862.6
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9314351.9
;      FILING DATE:  12-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9320988.0
;      FILING DATE:  12-OCT-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1669 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  ERT1B
US-08-522-421-1


  Query Match             2.4%;  Score 54.6;  DB 2;  Length 1669;
  Best Local Similarity   59.2%;  Pred. No. 1.1e-05;
  Matches   93;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy   1059 acttgttaagtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcctttat 1118
          | | || || ||| ||| ||||| | |||||| || ||   |     | ||| |
Db    965 AATAGTGAAATGGGCACCACAAAAACAGGTGCTTGCACATCCCGCGGTAGCAGGGTTTTT 1024

Qy   1119 cacccatgctggttcccatggtgtttatgaaagcatatgcaatggcgttcccatggtgat 1178
          ||| |||  |||||   ||  |     |||||| ||||| | | || || ||||||
Db   1025 CACTCATTGTGGTTGGAATTCTACGCTTGAAAGTATATGTGAAGAAGTCCCTATGGTGTG 1084

Qy   1179 gatgcccttgtttggtgatcagatggacaatgcaaag 1215
          | ||| ||  | |  || ||  |||  || |||| |
Db   1085 CAGGCCATTTCTAGCAGACCAACTGGTGAACGCAAGG 1121


RESULT  14
US-08-797-226-1
; Sequence 1, Application US/08797226
; Patent No. 5959180
;  GENERAL INFORMATION:
;    APPLICANT:  MOEHS, CHARLES P
;    APPLICANT:  ALLEN, PAUL V
;    APPLICANT:  ROCKHOLD, DAVID R
;    APPLICANT:  STAPLETON, ANDREW
;    APPLICANT:  GARBARINO, JOAN E
;    APPLICANT:  FRIEDMAN, MENDEL
```

```
Qy    560 cacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgcca 619
          ||||||  |   ||||   |||||| | || |   | ||          |||||   ||
Db    784 TACAGTG-TCAAATCCTTCCTCCTATATTCCTAAGTTACTAACGACCAATTCAGACCACA 842

Qy    620 tgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatgcc 679
          |||| |||  | | || ||   |||    ||   |   |||         |   | ||||
Db    843 TGACATTCCTGCAAAGGGTCAAGAACATGCTCTACCCTCTGGCCCTGTCCTACATTTGCC 902

Qy    680 accgttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttacgg 739
          |   |||||       |     |    | |||||||  || | || | |   ||  | |
Db    903 ATACTTTTTCTGCCCCTTATGCAAGTCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTCAG 962

Qy    740 agtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttggact 799
          |   |||||   |||| |  ||||  | ||| |||| |||  |||||||||  ||||||
Db    963 TGGTGGATCTTGTCAGCTATGCATCCGTGTGGCTGTTCCGAGGGGACTTTGTGATGGACT 1022

Qy    800 atcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcagggaa 859
          | ||||  ||  | ||||||||||| |||||||||| || ||||||||     | || |
Db   1023 ACCCCAGGCCGATCATGCCCAACATGGTCTTCATTGGGGGCATCAACTGTGCCAACGGGA 1082

Qy    860 agccgttgcctatggaattt 879
          ||||  |  ||  ||  | |
Db   1083 AGCCACTATCTCAGGTCTGT 1102
```

RESULT 11
PCT-US92-00282-8
; Sequence 8, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; TITLE OF INVENTION: THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1219 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US92-00282-8

Query Match 6.2%; Score 143.6; DB 5; Length 1219;
Best Local Similarity 50.8%; Pred. No. 2.3e-31;
Matches 419; Conservative 0; Mismatches 399; Indels 6; Gaps 3;

```
Qy     47 ttcctctatctgtgtgtctgctgacctgtggctttgccgaggcagggaagctactgg 106
          || ||  | | || |   |  |        || ||| | |  |  |  |  ||||| ||||
Db    245 TTTCTGCAGCGGTTTTCTTCTTAGCACTTTGGGGCATGGTTGTAGGTGACAAGCTGCTGG 304

Qy    107 tagtgcccatggatgggagccactggttcaccatgaggtcggtggtggagaaactcattc 166
          | || ||   ||| || |||||||||| | |  |||| |    | || |||   |||| |
Db    305 TGGTCCCTCAGGACGGAAGCCACTGGCTTAGTATGAAGGATATAGTTGAGGTTCTCAGTG 364

Qy    167 tcaggggcatgaggtggttgtagtcatgccagaggtgagttggcaactgggaagatcac 226
          | |||| |||||| | || || ||  |||| || || | || ||   ||  |  |||
Db    365 ACCGGGGTCATGAGATTGTAGTGGTGGTGCCTGAAGTTAATTTGCTTTTGAAAGAATCCA 424

Qy    227 tgaattgcacagtgaagacttattcaacttcatataccctggaggatctggaccgggagt 286
          | | ||||   || |  ||| ||    | |||  ||  || || ||| |       |
Db    425 AATACTACACAAGAAAAATCTATCCAGTGCCGTATGACCAAGAAGAGCTGAAGAACCGTT 484

Qy    287 tcaaggcttttgcccatgctcaatggaaagcacaagtacgaagtatatattctctattaa 346
          | | | ||||   |  ||| |    | | |  |     |    | |||    |
Db    485 ACCAATCATTTGGAAACAATCACTTTGCTGAGCGATCATTCCTAACTGCTCCTCAGACAG 544

Qy    347 tgggttcatacaatgacatt----tttgacttatttttttcaaattgcaggagtttgttt 402
          |       | ||    |||    || | || ||  ||    || |||  |||  |  |
Db    545 AGTACAGGAATAACATGATTGTTATTGGCCTGATACTTCATCAACTGCCAGAGCCTCCTG 604

Qy    403 aaagacaaaaaattagtagaatacttaaaggagagttcttttgatgcagtgtttctcgat 462
          | ||||   |     | | | ||| ||||||||   ||||||||  | ||   ||
Db    605 CAGGACAGGCACACCCTGAACTTCTTTAAGGAGAGCAAGTTTGATGCTCTTTTCACAGAC 664

Qy    463 ccttttgataactgtggcttaattgttgccaaatatttctccctccctccgtggtcttc 522
          ||          ||||||  | ||  | || |    ||   ||  ||          ||
Db    665 CCAGCCTTACCCTGTGGGGTGATCCTGGCGAGTATTTGGGCCTACCATCTGTGTACCTCT 724

Qy    523 gccaggggaatactttgccactatcttgaagaaggtgcacagtgccctgctcctctttcc 582
          |  |||   | | |   | ||       | |    ||| |   ||   | |    | |
Db    725 TCAGGGGTTTTCCGTGTTCCCTGGAGCATACATTCAGCAGAAGCCCAGACCCTGTGTCTC 784

Qy    583 tatgtccccagaattctcttagggttctcagatgccatgactttcaaggagagagtacgg 642
          |||||||       |  |  ||| |||||   |||||||||      |  ||||
Db    785 ACATTCCCCAGGTGCTACACAAAGTTTTCAGACCACATGACTTTTTCCCAACGAGTGGCC 844

Qy    643 aaccacatcatgcacttggaggaacatttattatgccaccgtttttttcaaaaatgcccta 702
          |||  | |  |  | |||  ||| | |  |   |||    |  || | ||   |||      |
Db    845 AACTTCCTTGTTAATTTGTTGGAGCCCTATCTATTTTATTGTCTGTTTTCAAAGTATGAA 904

Qy    703 gaaatagcctctgaaattctccaaacacctgtt-acggagtatgatctctacagccacac 761
          |||  | || || |   | ||| | | |  |||  |   | |    ||   |
Db    905 GAACTCGCATCAGCTGTCCTCAAGAGAGATGTGGATATAATCACCTTATATCATGAAGGT 964

Qy    762 atcaatttggttgttgcgaacggactttgttttggactatcccaaacccgtgatgcccaa 821
          ||  ||||| |||  ||   ||||||||  | || ||||| |  || || ||||||||
Db    965 CTCTGTTTGGCTGTTAAGATATGACTTTGTGCTTGAATATCCTAGGCCGGTCATGCCCAA 1024

Qy    822 catgatcttcattggtggtatcaactgccatcagggaaagccgt 865
          | || |||||||||| ||||||||||  |  ||  | ||| | |
Db   1025 C-TGGTCTTCATTGGAGGTATCAACTGTAAGAAGAGGAAGACTT 1067
```

RESULT 12
PCT-US92-00282-20
; Sequence 20, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; TITLE OF INVENTION: THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601

```
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-18


  Query Match             7.0%;  Score 161.8;  DB 5;  Length 1190;
  Best Local Similarity   51.8%;  Pred. No. 1.3e-36;
  Matches  414;  Conservative    0;  Mismatches  382;  Indels    3;  Gaps    2;

Qy      68 tgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatgggagcc 127
           ||||| |   |   | || | || |||||| ||||| |  | ||  ||||||| ||||
Db     140 TGCTGGGCCCAGTGGTGTCCCATGCTGGGAAGATACTGTTGATCCCAGTGGATGGCAGCC 199

Qy     128 actggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgaggtggttg 187
           ||||| | | ||||     ||   |  || | ||    | ||||| ||||  | ||||
Db     200 ACTGGCTGAGCATGCTTGGGGCCATCCAGCAGCTGCAGCAGAGGGGACATGAAATAGTTG 259

Qy     188 tagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaagactt 247
           |  |    || || |     | | |  |  ||     |     || |||  |||||||
Db     260 TCCTAGCACCTGACGCCTCGTTGTACATCAGAGACGGAGCATTTTACACCTTGAAGACGT 319

Qy     248 attcaacttcatataccctggaggatctggaccgggagtt--caaggcttttgcccatgc 305
           |  |     ||        ||||||| || |   |   ||    || ||   ||  |
Db     320 ACCCTGTGCCAATCCAAAGGGAGGATGTGAAAGAGTCTTTTGTTAGTCTCGGGCATAATG 379

Qy     306 tcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgacat 365
           |   ||  ||  |  | |         | |  || |   ||     | | ||  |
Db     380 TTTTTGAGAATGATTCTTTCCTGCAGCGTGTGATCAAAACATACAAGAAAATAAAAAAGG 439

Qy     366 ttttgact-tatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaat 424
             |   || |  |||| ||    ||       ||  |  |  ||||  |  |  | |  |
Db     440 ACTCTGCTATGCTTTTGTCTGGCTGTTCCCACTTACTGCACAACAAGGAGCTCATGGCCT 499

Qy     425 acttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggcttaa 484
           | |    || ||    |||||||   || |    || |||||   |   ||  ||   |
Db     500 CCCTGGCAGAAAGCAGCTTTGATGTCATGCTGACGGACCCTTTCCTTCCTTGCAGCCCCA 559

Qy     485 ttgttgccaaatatttctccctcccctccgtggtcttcgccagggaatactttgccact 544
           | || ||| | ||  | || || ||| | ||  |||||      | | | |  |||  |
Db     560 TCGTGGCCCAGTACCTGTCTCTGCCCACTGTATTCTTCTTGCATGCACTGCCATGCAGCC 619

Qy     545 atcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttag 604
               || | | ||||||||     ||  | ||||| || |||||   |||||
Db     620 TGGAATTTGAGGCTACCCAGTGCCCCAACCCATTCTCCTACGTGCCCAGGCCTCTCTCCT 679

Qy     605 ggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggagg 664
               ||||||  ||||||  |||  | || | ||   ||||    ||||   |||
Db     680 CTCATTCAGATCACATGACCTTCCTGCAGCGGGTGAAGAACATGCTCATTGCCTTTTCAC 739

Qy     665 aacatttattatgccaccgtttttttcaaaaatgccctagaaatagcctctgaaattctcc 724
           |  | ||  | ||| ||    |||              |  | ||||| ||| | || |
Db     740 AGAACTTTCTGTGCGACGTGGTTTATTCCCCGTATGCAACCCTTGCCTCAGAATTCCTTC 799

Qy     725 aaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacgg 784
           | | |   || || |   | || || |  |||    ||||  | ||| ||||  |||  |
Db     800 AGAGAGAGGTGACTGTCCAGGACCTATTGAGCTCTGCATCTGTCTGGCTGTTTAGAAGTG 859

Qy     785 actttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatca 844
           |||||||   |||  | || |  ||| | |||||||| ||| | ||  |||||| ||||
Db     860 ACTTTGTGAAGGATGACCCTAGGCCCATCATGCCCAATATGGTTTTTGTTGGTGGAATCA 919

Qy     845 actgccatcagggaaagcc 863
           |||||| |||   ||| ||
Db     920 ACTGCCTTCACCAAAATCC 938


RESULT  10
PCT-US92-00282-12
; Sequence 12, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1197 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-12


  Query Match             6.9%;  Score 161.2;  DB 5;  Length 1197;
  Best Local Similarity   52.6%;  Pred. No. 2e-36;
  Matches  421;  Conservative    0;  Mismatches  373;  Indels    6;  Gaps    3;

Qy      85 gccgaggcagggaagctactggtagtgcccatggatgggagccactgg--ttcaccatga 142
           || |||   || ||| |  |||| |||||||  |||||  |||| ||||  |    ||||
Db     304 GCTGAGAGTGGAAAGGTGTTGGTGGTGCCCACTGATGGCAGCCCCTGGTCTCAGTCATGC 363

Qy     143 ggtcggtggtggagaaactcattctcaggggcatgaggtggttgtagtcatgccagagg 202
           ||  ||   ||  | | |||  |  ||| || ||  ||| ||| ||  |||  |||||||
Db     364 GGGAGGCCTTGCGGGAGCTCCATGCCAGAGGCCACCAGGCGGTGGTCCTCACCCCAGAGG 423

Qy     203 tgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttcatata 262
           |||  |  |||  |   |  |         || |||  |||   | ||| |   | |||  |
Db     424 TGAATATGCACATCAAAGAAGAGAAATTTTTCACCCTGACAGCCTATGCTGTTCCATGGA 483

Qy     263 ccctggaggatctggaccgggagttcaaggcttttgcccatgctcaat---ggaaagcac 319
           |||  |  ||||   | || || |       ||  |    | || |     |     | || ||
Db     484 CCCAGAAGGAATTTGATCGCGTTACGCTGGGCTACACTCAAGGGTTCTTTGAAACAGAAC 543

Qy     320 aagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttatttt 379
           |  |  |  |  | |||||||  |   ||||        ||  |  | | |       ||
Db     544 ATCTTCTGAAGAGATATTCTAGAAGTATGGCAATTATGAACAATGTATCTTTGGCCCTTC 603

Qy     380 tttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagagtt 439
           |     |||    |  |  |  |  |      |  |     || | || |  | ||
Db     604 ATAGGTGTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGGCACCTGAATGCTACTT 663

Qy     440 cttttgatgcagtgtttctcgatcctttttgataactgtggcttaattgttgccaaatatt 499
           | |||||||  || ||    || ||  || |   ||||||      |  | || || ||
Db     664 CCTTTGATGTGGTTTTAACAGACCCCGTTAACCTCTGTGGGGCGGTGCTGGCTAAGTACC 723

Qy     500 tctccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaaggtg 559
           | ||  | ||  | ||| | ||    |||   || |  ||  |||       | |
Db     724 TGTCGATTCCTGCTGTGTTTTTTTGGAGGTACATTCCATGTGACTTAGACTTTAAGGCCA 783
```

```
Qy    858 aaagccgttgcctatgg 874
          |||||  |  ||  ||
Db    957 GAAGCCACTATCTCAGG 973


RESULT   8
PCT-US92-00282-14
; Sequence 14, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1066 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-14


  Query Match             7.9%;  Score 183.6;  DB 5;  Length 1066;
  Best Local Similarity   53.5%;  Pred. No. 6.7e-43;
  Matches  429;  Conservative    0;  Mismatches  369;  Indels    4;  Gaps    2;

Qy     81 ctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcaccat 140
          ||  || |||   || ||| |  |||| ||||||||  |||| |||||||||| ||| |||
Db    202 CTGGGCTGAGAGTGGAAAGGTGTTGGTGGTGCCCATTCATGGCAGCCACTGGCTCAGCAT 261

Qy    141 gaggtcggtggtggagaaactcattctcaggggcatgaggtggttgtagtcatgccaga 200
          | ||  |||  ||  | | |||  |  ||| || ||  |||  || ||  |||  |||||
Db    262 GCGGGAGGTCTTGCGGGAGCTCCATGCCAGAGGCCACCAGGCAGTGGTCCTCACCCCAGA 321

Qy    201 ggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttcata 260
          ||||| |  |||  |   |  |        || |||  |||  || ||| | | ||| |
Db    322 GGTGAATATGCACATCAAAGAAGAGAACTTTTTCACCCTGACAACCTATGCCATTTCGTG 381

Qy    261 taccctggaggatctggaccgggagttcaaggcttttgcccatgctcaat---ggaaagc 317
          |||| ||| ||  | || ||  |  |   ||      | ||     | |     | ||
Db    382 GACCCAGGATGAATTTGATCGCCATGTGCTGGGCCACACTCAACTGTACTTTGAAACAGA 441

Qy    318 acaactacgaagtatatattctctattaatgggttcatacaatgacatttttgacttatt 377
          |||  | |  |  | || ||    |  ||||     |  ||  | || | |    | |
Db    442 ACATTTTCTGAAGAAATTTTTCAGAAGTATGGCAATGTTGAACAATATGTCTTTGGTCTA 501

Qy    378 tttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagag 437
          |  |     |||   |   |  |  |  |  |     |  |     || | || |  |
Db    502 TCATAGGTCTTGTGTGGAGCTACTACATAATGAGGCCCTGATCAGGCACCTGAATGCTAC 561

Qy    438 ttcttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaata 497
          ||| |||||||  || ||    || ||  || |   ||| |    | |  | || || ||
Db    562 TTCCTTTGATGTGGTTTTAACAGACCCCGTTAACCTCTGCGCGGCAGTGCTGGCTAAGTA 621

Qy    498 tttctccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaagg 557
          | ||  | ||  | ||| | ||    |||   || |  ||  | |       | ||
Db    622 CCTGTCGATTCCTACTGTGTTTTTTTTGAGGAACATTCCATGTGATTTAGACTTTAAGGG 681

Qy    558 tgcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgc 617
          |||||||  ||    |||   ||||||  | ||  ||| | ||    |      |||||
Db    682 CACACAGTGTCCAAACCCTTCCTCCTATATTCCTAGATTACTAACAACCAATTCAGACCA 741

Qy    618 catgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatg 677
          ||||||  |||| | | || ||   ||||    ||   |   |||        |   | ||
Db    742 CATGACATTCATGCAAAGGGTCAAGAACATGCTCTACCCTCTGGCCCTGTCCTACATTTG 801

Qy    678 ccaccgtttcttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttac 737
          |||   |||||       |           | ||||||||  || | || | |   ||  |
Db    802 CCATGCTTTTTCTGCTCCTTATGCCAGCCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTC 861

Qy    738 ggagtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttgga 797
          | |   ||| |   ||| ||  ||||  | ||| |||| |||  |||||||||  ||||
Db    862 AGTGGTGGATATTCTCAGTCATGCATCTGTGTGGCTGTTCCGAGGGGACTTTGTGATGGA 921

Qy    798 ctatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggg 857
          ||| ||||  ||  | ||||||||||| ||||||||| || |||||||     |  |
Db    922 CTACCCCAGGCCAATCATGCCCAACATGGTCTTCATTGG-GGCATCAACTGTGCCAACAG 980

Qy    858 aaagccgttgcctatggaattt 879
          |||||  |  ||  ||  | |
Db    981 GAAGCCACTATCTCAGGTCTGT 1002


RESULT   9
PCT-US92-00282-18
; Sequence 18, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1190 base pairs
```

```
Db    303 GGTGAATATGTACATCAAAGAAGAGAACTTTTTCACCCTGACAACGTATGCCATTTCATG 362

Qy    261 taccctggaggatctggaccgggagttcaaggcttttgcccatgctcaat---ggaaagc 317
          |||| ||| ||  | || ||    ||   || |    | ||  |    |     | ||
Db    363 GACCCAGGACGAATTTGATCGCCTTTTGCTGGGTCACACTCAATCGTTCTTTGAAACAGA 422

Qy    318 acaagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttatt 377
          |||  | |  |  | || ||||  |  ||||        ||  | || | |    |  |
Db    423 ACATCTTCTGATGAAATTTTCTAGAAGAATGGCAATTATGAACAATATGTCTTTGATCAT 482

Qy    378 tttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagag 437
             |     |||   |    |  |  |  |  |     |  |     || |  | |  |
Db    483 ACATAGGTCTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGGCACCTGCATGCTAC 542

Qy    438 ttcttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaata 497
          |||  |||||||  ||  |    || || ||| |   |||  |     |  | || || ||
Db    543 TTCCTTTGATGTGGTTCTAACAGACCCCTTTCACCTCTGCGCGGCGGTGCTGGCTAAGTA 602

Qy    498 tttctccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaagg 557
            | ||  | ||  | ||| | |||   |||   || |  ||  | |          | ||
Db    603 CCTGTCGATTCCTGCTGTGTTTTTCTTGAGGAACATTCCATGTGATTTAGACTTTAAGGG 662

Qy    558 tgcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgc 617
            |||||||  ||    |||   |||||| | || ||| | ||          |||||
Db    663 CACACAGTGTCCAAACCCTTCCTCCTATATTCCTAGATTACTAACGACCAATTCAGACCA 722

Qy    618 catgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatg 677
          ||||||  |||  | | || ||   ||||    ||   |   |||         |   | ||
Db    723 CATGACATTCCTGCAAAGGGTCAAGAACATGCTCTACCCTCTGGCCCTGTCCTACCTTTG 782

Qy    678 ccaccgttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttac 737
          |||   | |||       |     |    | ||||||||  || | || | |   ||  |
Db    783 CCATGCTGTTTCTGCTCCTTATGCAAGCCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTC 842

Qy    738 ggagtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttgga 797
            | |   |||||   ||||||  ||||    ||| |||| |||  ||| |||||  ||||
Db    843 AGTGGTGGATCTTGTCAGCCATGCATCCTGGTGGCTGTTCCGAGGGGAGTTTGTGATGGA 902

Qy    798 ctatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggg 857
            || ||||  ||  | |||||||||||| ||||||||| || ||||||||     | ||
Db    903 TTACCCCAGGCCGATCATGCCCAACATGGTCTTCATTGGGGGCATCAACTGTGCCAACGG 962

Qy    858 aaagccgttgcctatggaatttgaagcctacatta 892
            |||||  |  ||  ||  | |   |    | |||
Db    963 GAAGCCACTATCTCAGGTCTGTATTGGTGCCTTTA 997


RESULT   7
PCT-US92-00282-16
; Sequence 16, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1448 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  YES
PCT-US92-00282-16


  Query Match             8.0%;  Score 186.6;  DB 5;  Length 1448;
  Best Local Similarity   53.8%;  Pred. No. 1.1e-43;
  Matches  429;  Conservative    0;  Mismatches  364;  Indels    4;  Gaps    2;

Qy     81 ctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcaccat 140
          ||  || |||   |||||| | ||||| || ||||| ||||| |||||||| ||  |||
Db    178 CTGGGCTGAGAGTGGGAAGGTGCTGGTGGTACCCATTGATGGCAGCCACTGGCTCG-CAT 236

Qy    141 gaggtcggtggtggagaaactcattctcaggggcatgaggtggttgtagtcatgccaga 200
          | ||  ||  |||  | | |||  |  ||| || ||  ||| ||| ||  |||  |||||
Db    237 GCGGGAGGCCGTGCGGGAGCTCCATGCCAGAGGCCACCAGGCGGTGGTCCTCACCCCAGA 296

Qy    201 ggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttcata 260
          ||||| |  |||  |   | |        || |||  |||  || ||| | | |||||
Db    297 GGTGAATATGCACATCAAACAAGAGAAATTTTTCACCCTGACAACCTATGCCATTTCATG 356

Qy    261 taccctggaggatctggaccgggagttcaaggcttttgcccatgctcaat---ggaaagc 317
          |||| ||| ||  | || ||    |   ||       | ||     | |    | ||
Db    357 GACCCAGGATGAATTTGATCGCCTTGTGCTGGGCCACACTCAACTGTACTTTGAAACAGA 416

Qy    318 acaagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttatt 377
          |||  | |  |  | || ||||  |   ||||     |  ||  |    | |    | ||
Db    417 ACATTTTCTGAAGACATTTTCTAAAAGTATGGCAATTTTGAAAAATTCATCTTTGGTCTT 476

Qy    378 tttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagag 437
          |  |     |||  |    |  |  |  |  |     |  |     || | || | |  |
Db    477 TCATAGGTCTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGGCACCTGAATGCTAC 536

Qy    438 ttcttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaata 497
          ||| || ||||  || ||  | || ||  || |   ||||||    | |  | || || ||
Db    537 TTCCTTCGATGTGGTTTTAATAGACCCAATTTACCTCTGTGGGGCAGTGCTGGCTAAGTA 596

Qy    498 tttctccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaagg 557
            | ||  | ||  | ||| | ||    |||   || |  ||  | |          | ||
Db    597 CCTATCAATTCCTGCTGTGTTTTTTTTGAGGAACATTCCATGTGATTTAGACTTTAAGGG 656

Qy    558 tgcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgc 617
            |||||||  ||   ||||   |||||| | || |   | ||          |||||
Db    657 CACACAGTGTCCAAATCCTTCCTCCTATATTCCTAAGTTACTAACGACCAATTCAGACCA 716

Qy    618 catgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatg 677
          ||||||  |||  | | || ||   ||||    ||   |   |||         |   | ||
Db    717 CATGACATTCCTGCAAAGGGTCAAGAACATGCTCTACCCTCTGGCCCTGTCCTACCTTTG 776

Qy    678 ccaccgttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttac 737
          |||   | |||       |     |    | ||||||||  || | || | |   ||  |
Db    777 CCATGCTCTTTCTGCTCCTTATGCAAGCCTTGCCTCTGAGCTTTTTCAGAGGGAGGTGTC 836

Qy    738 ggagtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttgga 797
            | |   |||||   ||||||  ||||  | ||| |||| |||  |||||||||  ||||
Db    837 AGTGGTGGATCTTCTCAGCCATGCATCTGTGTGGCTGTTCCGAGGGGACTTTGTGATGGA 896

Qy    798 ctatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggg 857
            || ||||  ||  | |||||||||||| |||||||||| || ||||||||     | |
Db    897 TTACCCCAGGCCGATCATGCCCAACATGGTCTTCATTGGGGGCATCAACTGTGCCAACAG 956
```

```
Qy   1251 tgttctggaaatgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaa 1310
          |   |   ||| |   |  ||||||    |||||  | ||  ||||||| ||||||
Db   1302 catcaggaccatgtcaagtagagatttgctcaatgcattgaagtcagtcattaatgaccc 1361

Qy   1311 aagttacaaggagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagcc 1370
          |  || || ||||| ||||||    | || ||  ||||  | || |  |||||| ||||
Db   1362 tatctataaagagaatatcatgaaattatcaagaattcatcatgatcaaccggtgaagcc 1421

Qy   1371 gctggacctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacct 1430
          ||||| |  || || |||||| | ||||||| ||| | || || || ||    |||||
Db   1422 cctggatcgagcagtcttctggattgagtttgtcatgcgccataaaggagccaagcacct 1481

Qy   1431 gcgccccgcagcccacgacctcacctggtaccagtaccattccttggacgtgattggttt 1490
          ||   |||||||||| ||||||||||  |||||||| || ||||| ||||| |  ||
Db   1482 tcgggtcgcagcccacaacctcacctggatccagtaccactctttggatgtgatagcatt 1541

Qy   1491 cctcttggccgtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggcta 1550
          |||  |||||  ||||   ||  ||   || || | |   ||||||||    | | | |
Db   1542 cctgctggcctgcgtggcaactatgatatttatgatcacaaaatgttgcctgttttgttt 1601

Qy   1551 ccggaa 1556
          ||| ||
Db   1602 ccgaaa 1607


RESULT   5
PCT-US92-00282-22
; Sequence 22, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  783 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-22


  Query Match             9.6%;  Score 222.4;  DB 5;  Length 783;
  Best Local Similarity   99.6%;  Pred. No. 3.9e-54;
  Matches  223;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1093 ggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgtttatgaaagc 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    471 GGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAGC 530

Qy   1153 atatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatggacaatgca 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    531 ATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAATGCA 590

Qy   1213 aagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatgacttctgaa 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    591 AAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGAA 650

Qy   1273 gatttagaaaatgctctaaaagcagtcatcaatgacaaaagtta 1316
          |||||||||||||||||||||||||||||||||||||||||| ||
Db    651 GATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGGTA 694


RESULT   6
PCT-US92-00282-10
; Sequence 10, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1008 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-10


  Query Match             8.3%;  Score 192.6;  DB 5;  Length 1008;
  Best Local Similarity   53.1%;  Pred. No. 1.7e-45;
  Matches  433;  Conservative    0;  Mismatches  379;  Indels    3;  Gaps    1;

Qy     81 ctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcaccat 140
          ||  || |||   |||||| | ||||| |||||||  ||||| |||||||| ||| |||
Db    183 CTGGGCTGAGAGTGGGAAGGTGCTGGTGGTGCCCACTGATGGCAGCCACTGGCTCAGCAT 242

Qy    141 gaggtcggtggtggagaaactcattctcaggggggcatgaggtggttgtagtcatgccaga 200
          | ||  ||   ||  | | |||  |   || || ||  ||||||| ||  |||  |  ||
Db    243 GCGGGAGGCCTTGCGGGACCTCCATGCGAGAGGCCACCAGGTGGTGGTCCTCACCCTGGA 302

Qy    201 ggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttcata 260
          ||||| |  | |  |   |  |         || |||  |||  || ||| | | |||||
```

```
Db    847 TTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGAGTGCGGGATTCAAAGGTGGTCCCA 906

Qy   1949 ccgctgcccctactgcaaatggcagttttaatcttatcttttggcttctgcagatggttg 2008
Db    907 CGGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTTG 966

Qy   2009 caattgatccttaaccaataatggtcagtcctcatctctgtcctgcttcataggtgccac 2068
Db    967 CAATTGATCCTTAACCAATAATGGTCAGTCCTCATCTCTGTCGTGCTTCATAGGTGCCAC 1026

Qy   2069 cttgtgtgtttaaagaagggaagctttgtacctttagagtgtaggtgaaatgaatgaatg 2128
Db   1027 CTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATG 1086

Qy   2129 gcttggagtgcactgagaacagcatatgatttcttgctttggggaaaaagaatgatgcta 2188
Db   1087 GCTTGGAGTGCACTGAGAACAGCATATGATTTCTTGCTTTGGGGAAAAAGAATGATGCTA 1146

Qy   2189 tgaaattggtgggtggtgtatttgagaagataatcattgcttatgtcaaatggagctgaa 2248
Db   1147 TGAAATTGGTGGGTGGTGTATTTGAGAAGATAATCATTGCTTATGTCAAATGGAGCTGAA 1206

Qy   2249 tttgataaaaacccaaaatacagctatgaagtgctgggcaagtttacttttttctgatg 2308
Db   1207 TTTGATAAAAACCCAAAATACAGCTATGAAGTGCTGGGCAAGTTTACTTTTTTTCTGATG 1266

Qy   2309 tttcctacaact 2320
Db   1267 TTTCCTACAACT 1278
```

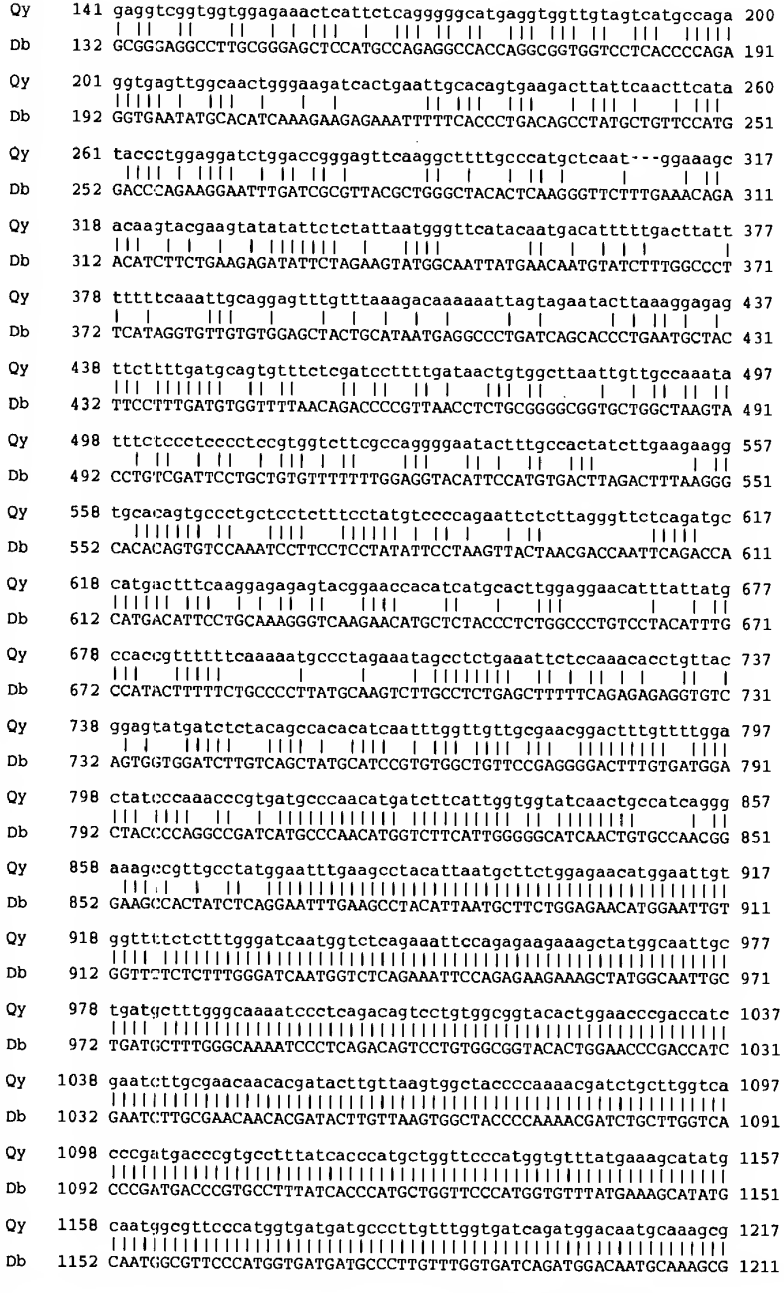
RESULT 4
US-09-180-852-1
; Sequence 1, Application US/09180852
; Patent No. 6287834
; GENERAL INFORMATION:
; APPLICANT: BELANGER, Alain
; APPLICANT: HUM, Dean W.
; APPLICANT: BEAULIEU, Martin
; APPLICANT: LEVESQUE, Eric
; TITLE OF INVENTION: CHARACTERIZATION AND USE OF AN ISOLATED URIDINE
; TITLE OF INVENTION: DIPHOSPHO-GLUCURONOSYLTRANSFERASE
; FILE REFERENCE: 1259-449
; CURRENT APPLICATION NUMBER: US/09/180,852
; CURRENT FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: PCT/CA97/00328
; EARLIER FILING DATE: 1997-05-16
; EARLIER APPLICATION NUMBER: US 08/649,319
; EARLIER FILING DATE: 1996-05-17
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2107
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52)..(1644)
US-09-180-852-1

Query Match 14.8%; Score 344.4; DB 4; Length 2107;
Best Local Similarity 53.9%; Pred. No. 5.6e-89;
Matches 811; Conservative 0; Mismatches 671; Indels 24; Gaps 4;

```
Qy     75 ctgtggctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggtt 134
Db    102 ctttagctctgggagttgtggaaaggtgctggtgtggcccacagaatacagccattggat 161

Qy    135 caccatgaggtcggtggtggagaaactcattctcaggggcatgaggtggttgtagtcat 194
Db    162 aaatatgaagacaatcctggaagagcttgttcagaggggtcatgaggtgattgtgttgac 221

Qy    195 gccagaggtgagttggcaactgg------gaagatcactgaattgcacagtgaagactta 248
Db    222 atcttcggcttctattcttgtcaatgccagtaaatcatctgctattaaattagaagttta 281

Qy    249 ttcaacttcatataccctggaggatctggaccggga-----------gttcaaggcttt 296
Db    282 tcctacatctttaactaaaaatgatttggaagatttttttatgaaaatgttcgatagatg 341

Qy    297 tgcccatgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcata 356
Db    342 gacatatagtatttcaaaaaatacattttggtcatatttttcacaactacaagaattgtg 401

Qy    357 caatgacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaatt 416
Db    402 ttgggaatattctgactataatataaagctctgtgaagatgcagttttgaacaagaaact 461

Qy    417 agtagaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactg 476
Db    462 tatgagaaaactacaagagtcaaaatttgatgtccttctggcagatgccgttaatccctg 521

Qy    477 tggcttaattgttgccaaatatttctccctccccctccgtggtcttcgccaggggaatact 536
Db    522 tggtgagctgctggctgaactacttaacataccctttctgtacagtctccgcttctctgt 581

Qy    537 ttgccactatcttg---aagaaggtgcacagtgccctgctcctctttcctatgtccccag 593
Db    582 tggctacacagttgagaagaatggtggaggatttctgttccctccttcctatgtacctgt 641

Qy    594 aattctcttagggttctcagatgccatgactttcaaggagagagtacggaaccacat--- 650
Db    642 tgttatgtcagaattaagtgatcaaatgattttcatggagaggataaaaaatatgatata 701

Qy    651 catgcacttggaggaacatttattatgccaccgtttttttcaaaaatgccctagaaatagc 710
Db    702 tatgctttattttgacttttggtttcaagcatatgatctgaagaagtgggaccagtttta 761

Qy    711 ctctgaaattctccaaacacctgttacggagtatgatctctacagccacacatcaatttg 770
Db    762 tagtgaagttctaggaagacccactacattatttgagacaatggggaaagctgaaatgtg 821

Qy    771 gttgttgcgaacggactttgttttggactatcccaaacccgtgatgcccaacatgatctt 830
Db    822 gctcattcgaacctattgggatttgaatttcctcgcccattcttaccaaatgttgattt 881

Qy    831 cattggtggtatcaactgccatcagggaaagccgttgcctatggaatttgaagcctacat 890
Db    882 tgttggaggacttcactgtaaaccagccaaacccttgcctaaggaaatggaagagtttgt 941

Qy    891 taatgcttctggagaacatggaattgtggttttctctttgggatcaatggtctcagaaat 950
Db    942 gcagagctctggagaaaatggtattgtggtgttttctctggggtcgatgatcagtaacat 1001

Qy    951 tccagagaagaaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcct 1010
Db   1002 gtcagaagaaagtgccaacatgattgcatcagcccttgcccagatcccacaaaaggttct 1061

Qy   1011 gtggcggtacactggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtg 1070
Db   1062 atggagatttgatggcaagaagccaaatactttaggttccaatactcgactgtataagtg 1121

Qy   1071 gctaccccaaaacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctgg 1130
Db   1122 gttaccccagaatgaccttcttggtcatcccaaaaccaaagctttataactcatggtgg 1181

Qy   1131 ttcccatggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtt 1190
Db   1182 aaccaatggcatctatgaggcgatctaccatgggatccctatggtgggcattcccttgtt 1241

Qy   1191 tggtgatcagatggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaa 1250
Db   1242 tgcggatcaacatgataacattgctcacatgaaagccaagggagcagccctcagtgtgga 1301
```

```
Qy   1505 tgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttgg 1564
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1496 TGCTGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGG 1555

Qy   1565 ggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtggga 1624
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1556 GGAAAAAAGGGCGAGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGA 1615

Qy   1625 aataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtg 1684
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1616 AATAAGGTAAAATTTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTG 1675

Qy   1685 ttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtg 1744
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1676 TTAAATTCATTTTATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTG 1735

Qy   1745 ctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaat 1804
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1736 CTTTAAAAAATTCTCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAAT 1795

Qy   1805 atgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggac 1864
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1796 ATGTATCGTGCCCCCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGAC 1855

Qy   1865 gtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttggga 1924
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1856 GTGCAGACAGGCTGGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGA 1915

Qy   1925 gtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatctta 1984
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1916 GTGCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTA 1975

Qy   1985 tcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatc 2044
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1976 TCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCTCATC 2035

Qy   2045 tctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtaccttta 2104
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2036 TCTGTCCTGCTTCATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTTA 2095

Qy   2105 gagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttcttg 2164
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2096 GAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTTCTTG 2155

Qy   2165 ctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatca 2224
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2156 CTTTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATAATCA 2215

Qy   2225 ttgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgctg 2284
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2216 TTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGTGCTG 2275

Qy   2285 ggcaagtttacttttttttctgatgtttcctacaact 2320
          |||| ||||||||||||||||||||||||||||||||
Db   2276 GGCAAGTTTACTTTTTTTCTGATGTTTCCTACAACT 2311


RESULT   3
PCT-US92-00282-25
; Sequence 25, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714527 CUSH
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1561 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-25


  Query Match             41.7%;  Score 967.2;  DB 5;  Length 1561;
  Best Local Similarity   98.0%;  Pred. No. 3.3e-268;
  Matches 1011;  Conservative    0;  Mismatches   18;  Indels    3;  Gaps    3;

Qy   1291 aaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctccagccttcac 1350
          || | |    ||  ||     ||||||||||||||||||||||||||||||||||||||||
Db    248 AACGAAACTGTCTTTGTGTTTAGTTACAAGGAGAACATCATGCGCCTCTCCAGCCTTCAC 307

Qy   1351 aaggaccgccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgagg 1410
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    308 AAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGG 367

Qy   1411 cacaagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccat 1470
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    368 CACAAGGGCGCGCCACACCTGCGCCCCGCAGCCCACGACCTCACCTGGTACCAGTACCAT 427

Qy   1471 tccttggacgtgattggtttcctcttggccgtcgtgct-gacagtggccttcatcacctt 1529
          |||||||||||||||||||||||||||||||||||||  ||| ||||||||||||||||||
Db    428 TCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCTGGCTGGACAGTGGCCTTCATCACCTT 487

Qy   1530 taaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcgagttaagaaagc 1589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    488 TAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGGAAAAAAGGGCGAGTTAAGAAAGC 547

Qy   1590 ccacaaatccaagacccattgagaagtgggtgggaaataaggtaaaattttgaaccattc 1649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    548 CCACAAATCCAAGACCCATTGAGAAGTGGGTGGGAAATAAGGTAAAATTTTGAACCATTC 607

Qy   1650 cctagtcatttccaaacttgaaaacagaatcagtgttaaattcattttattcttattaag 1709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    608 CCTAGTCATTTCCAAACTTGAAAACAGAATCAGTGTTAAATTCATTTTATTCTTATTAAG 667

Qy   1710 gaaatactttgcataaattaatcagccccagagtgctttaaaaaattctcttaaataaaa 1769
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    668 GAAATACTTTGCATAAATTAATCAGCCCCAGAGTGCTTTAAAAAATTCTCTTAAATAAAA 727

Qy   1770 ataatagactcgctagtcagtaaagatatttgaatatgtatcgtg-cccctccggtgtc 1828
          |||||||||||||||||||||||||||||||||||||||||||| ||||||| ||||||
Db    728 ATAATAGACTCGCTAGTCAGTAAAGATATTTGAATATGTATCGTGCCCCCCTCTGGTGTC 787

Qy   1829 tttgatcaggatgacatgtgccatttttcagaggacgtgcagacaggctggcattctaga 1888
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    788 TTTGATCAGGATGACATGTGCCATTTTTCAGAGGACGTGCAGACAGGCTGG-ATTCTAGA 846

Qy   1889 ttacttttcttactctgaaacatggcctgtttgggagtgcgggattcaaaggtggtccca 1948
```

```
          ||||||||||||||||||||||
Db   2292 TTTTTCTGATGTTTCCTACAACT 2314


RESULT   2
PCT-US92-00282-1
; Sequence 1, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2336 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-1


  Query Match             69.4%;  Score 1609.2;  DB 5;  Length 2336;
  Best Local Similarity   82.7%;  Pred. No. 0;
  Matches 1865;  Conservative    0;  Mismatches  388;  Indels    3;  Gaps    2;

Qy      68 tgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatgggagcc 127
           |||||  |   |   |  || | || ||||||  |||||  |  | ||  |||||||  ||||
Db      56 TGCTGGGCCCAGTGGTGTCCCATGCTGGGAAGATACTGTTGATCCCAGTGGATGGCAGCC 115

Qy     128 actggttcaccatgaggtcggtggtggagaaactcattctcagggggcatgaggtggttg 187
           |||||  | |  ||||     ||   |  || | ||    |  |||||  |||||  | ||||
Db     116 ACTGGCTGAGCATGCTTGGGGCCATCCAGCAGCTGCAGCAGAGGGGACATGAAATAGTTG 175

Qy     188 tagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaagactt 247
           |  |    || || |     | | |  |  ||     |     || |||  |||||||  |
Db     176 TCCTAGCACCTGACGCCTCGTTGTACATCAGAGACGGAGCATTTTACACCTTGAAGACGT 235

Qy     248 attcaacttcatataccctggaggatctggaccgggagtt--caaggcttttgcccatgc 305
           |  |     |||       ||||||| || |   |   ||    || ||   ||  |
Db     236 ACCCTGTGCCATTCCAAAGGGAGGATGTGAAAGAGTCTTTTGTTAGTCTCGGGCATAATG 295

Qy     306 tcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgacat 365
           |   ||  ||  |    | |        | |  || |   ||      | | ||  |
Db     296 TTTTTGAGAATGATTCTTTCCTGCAGCGTGTGATCAAAACATACAAGAAAATAAAAAAGG 355

Qy     366 ttttgact-tatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaat 424
             |   || |  |||| ||    ||       ||  |  |  ||||  |  |  | |  |

Db     356 ACTCTGCTATGCTTTTGTCTGGCTGTTCCCACTTACTGCACAACAAGGAGCTCATGGCCT 415

Qy     425 acttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggcttaa 484
           | |    || ||    |||||||  || |    || |||||   |   ||  ||   |
Db     416 CCCTGGCAGAAAGCAGCTTTGATGTCATGCTGACGGACCCTTTCCTTCCTTGCAGCCCCA 475

Qy     485 ttgttgccaaaatatttctccctcccctccgtggtcttcgccaggggaatactttgccact 544
           | || ||| | ||  | || || ||| | ||  |||||       | | | |  |||  |
Db     476 TCGTGGCCCAGTACCTGTCTCTGCCCACTGTATTCTTCTTGCATGCACTGCCATGCAGCC 535

Qy     545 atcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttag 604
                 || | | | ||||||||    ||  | |||||  || |||||   |||||
Db     536 TGGAATTTGAGGCTACCCAGTGCCCCAACCCATTCTCCTACGTGCCCAGGCCTCTCTCCT 595

Qy     605 ggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggagg 664
               ||||||  ||||||  |||  | || | ||   ||||    ||||   |||
Db     596 CTCATTCAGATCACATGACCTTCCTGCAGCGGGTGAAGAACATGCTCATTGCCTTTTCAC 655

Qy     665 aacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaattctcc 724
           |   | ||   | ||| ||    |||             |    | |||||  ||| | || |
Db     656 AGAACTTTCTGTGCGACGTGGTTTATTCCCCGTATGCAACCCTTGCCTCAGAATTCCTTC 715

Qy     725 aaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacgg 784
           | | |   || || |   | || || |  |||    ||||  | ||| ||||  |||  |
Db     716 AGAGAGAGGTGACTGTCCAGGACCTATTGAGCTCTGCATCTGTCTGGCTGTTTAGAAGTG 775

Qy     785 actttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatca 844
           |||||||   ||| || || |  ||| | |||||||| ||| | ||  |||||| ||||
Db     776 ACTTTGTGAAGGATTACCCTAGGCCCATCATGCCCAATATGGTTTTTGTTGGTGGAATCA 835

Qy     845 actgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctggag 904
           ||||||  |||   |||  ||  |  |   ||||||||||||||||||||||||||||||
Db     836 ACTGCCTTCACCAAAATCCACTATCCCAGGAATTTGAAGCCTACATTAATGCTTCTGGAG 895

Qy     905 aacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaagaaag 964
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     896 AACATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAG 955

Qy     965 ctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactg 1024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     956 CTATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTG 1015

Qy    1025 gaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacg 1084
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1016 GAACCCGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACG 1075

Qy    1085 atctgcttggtcacccgatgacccgtgccttttatcacccatgctggttcccatggtgttt 1144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1076 ATCTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTT 1135

Qy    1145 atgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatgg 1204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1136 ATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGG 1195

Qy    1205 acaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatga 1264
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1196 ACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGA 1255

Qy    1265 cttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggaga 1324
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1256 CTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGA 1315

Qy    1325 acatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccg 1384
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1316 ACATCAGGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCG 1375

Qy    1385 tgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagccc 1444
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1376 TGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCC 1435

Qy    1445 acgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcg 1504
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1436 ACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCG 1495
```


Qy 141 gaggtcggtggtggagaaactcattctcaggggcatgaggtggttgtagtcatgccaga 200
Db 132 GCGGGAGGCCTTGCGGGAGCTCCATGCCAGAGGCCACCAGGCGGTGGTCCTCACCCCAGA 191
Qy 201 ggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttcata 260
Db 192 GGTGAATATGCACATCAAAGAAGAGAAATTTTTCACCCTGACAGCCTATGCTGTTCCATG 251
Qy 261 taccctggaggatctggaccgggagttcaaggcttttgcccatgctcaat---ggaaagc 317
Db 252 GACCCAGAAGGAATTTGATCGCGTTACGCTGGGCTACACTCAAGGGTTCTTTGAAACAGA 311
Qy 318 acaagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttatt 377
Db 312 ACATCTTCTGAAGAGATATTCTAGAAGTATGGCAATTATGAACAATGTATCTTTGGCCCT 371
Qy 378 tttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagag 437
Db 372 TCATAGGTGTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGCACCCTGAATGCTAC 431
Qy 438 ttcttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaata 497
Db 432 TTCCTTTGATGTGGTTTTAACAGACCCCGTTAACCTCTGCGGGGCGGTGCTGGCTAAGTA 491
Qy 498 tttctccctccccctccgtggtcttcgccaggggaatactttgccactatcttgaagaagg 557
Db 492 CCTGTCGATTCCTGCTGTGTTTTTTTGGAGGTACATTCCATGTGACTTAGACTTTAAGGG 551
Qy 558 tgcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgc 617
Db 552 CACACAGTGTCCAAATCCTTCCTCCTATATTCCTAAGTTACTAACGACCAATTCAGACCA 611
Qy 618 catgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatg 677
Db 612 CATGACATTCCTGCAAAGGGTCAAGAACATGCTCTACCCTCTGGCCCTGTCCTACATTTG 671
Qy 678 ccaccgtttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttac 737
Db 672 CCATACTTTTTCTGCCCCTTATGCAAGTCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTC 731
Qy 738 ggagtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttgga 797
Db 732 AGTGGTGGATCTTGTCAGCTATGCATCCGTGTGGCTGTTCCGAGGGGACTTTGTGATGGA 791
Qy 798 ctatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggg 857
Db 792 CTACCCCAGGCCGATCATGCCCAACATGGTCTTCATTGGGGGCATCAACTGTGCCAACGG 851
Qy 858 aaagccgttgcctatggaatttgaagcctacattaatgcttctggagaacatggaattgt 917
Db 852 GAAGCCACTATCTCAGGAATTTGAAGCCTACATTAATGCTTCTGGAGAACATGGAATTGT 911
Qy 918 ggttttctctttgggatcaatggtctcagaaattccagagaagaaagctatggcaattgc 977
Db 912 GGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATGGCAATTGC 971
Qy 978 tgatgctttgggcaaaatccctcagacagtcctgtggcggtacactggaacccgaccatc 1037
Db 972 TGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGGAACCCGACCATC 1031
Qy 1038 gaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacgatctgcttggtca 1097
Db 1032 GAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACGATCTGCTTGGTCA 1091
Qy 1098 cccgatgacccgtgccttttatcacccatgctggttcccatggtgtttatgaaagcatatg 1157
Db 1092 CCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAGCATATG 1151
Qy 1158 caatggcgttcccatggtgatgatgcccttgtttggtgatcagatggacaatgcaaagcg 1217
Db 1152 CAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAATGCAAAGCG 1211
Qy 1218 catggagactaagggagctggagtgaccctgaatgttctggaaatgacttctgaagattt 1277
Db 1212 CATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGAAGATTT 1271
Qy 1278 agaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcct 1337
Db 1272 AGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAACATCAGGCGCCT 1331
Qy 1338 ctccagccttcacaaggaccgcccggtggagccgctggacctggccgtgttctgggtgga 1397
Db 1332 CTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGA 1391
Qy 1398 gtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagcccacgacctcacctg 1457
Db 1392 GTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCCACGACCTCACCTG 1451
Qy 1458 gtaccagtaccattccttggacgtgattggtttcctcttggccgtcgtgctgacagtggc 1517
Db 1452 GTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGC 1511
Qy 1518 cttcatcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcg 1577
Db 1512 CTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAATGCTTGGGGAAAAAAGGGCG 1571
Qy 1578 agttaagaaagcccacaaatccaagacccattgagaagtgggtgggaaataaggtaaaat 1637
Db 1572 AGTTAAGAAAGCCCACAAATCCAAGACCCATTGAGAAGTGGGTGGGAAATAAGGTAAAAT 1631
Qy 1638 tttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgttaaattcatttt 1697
Db 1632 TTTGAACCATTCCCTAGTCATTTCCAAACTTGAAAACAGAATCAGTGTTAAATTCATTTT 1691
Qy 1698 attcttattaaggaaatactttgcataaattaatcagccccagagtgctttaaaaaattc 1757
Db 1692 ATTCTTATTAAGGAAATACTTTGCATAAATTAATCAGCCCCAGAGTGCTTTAAAAAATTC 1751
Qy 1758 tcttaaataaaaataatagactcgctagtcagtaaagatatttgaatatgtatcgtgccc 1817
Db 1752 TCTTAAATAAAAATAATAGACTCGCTAGTCAGTAAAGATATTTGAATATGTATCGTGCCC 1811
Qy 1818 cctccggtgtctttgatcaggatgacatgtgccatttttcagaggacgtgcagacaggct 1877
Db 1812 CCTCCGGTGTCTTTGATCAGGATGACATGTGCCATTTTTCAGAGGACGTGCAGACAGGCT 1871
Qy 1878 ggcattctagattacttttcttactctgaaacatggcctgtttgggagtgcgggattcaa 1937
Db 1872 GGCATTCTAGATTACTTTTCTTACTCTGAAACATGGCCTGTTTGGGAGTGCGGGATTCAA 1931
Qy 1938 aggtggtcccaccgctgcccctactgcaaatggcagttttaatcttatcttttggcttct 1997
Db 1932 AGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTATCTTTTGGCTTCT 1991
Qy 1998 gcagatggttgcaattgatccttaaccaataatggtcagtcctcatctctgtcctgcttc 2057
Db 1992 GCAGATGGTTGCAATTGATCCTTAACCAATAATGGTCAGTCCTCATCTCTGTCCTGCTTC 2051
Qy 2058 ataggtgccaccttgtgtgtttaaagaagggaagctttgtaccttagagtgtaggtgaa 2117
Db 2052 ATAGGTGCCACCTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAA 2111
Qy 2118 atgaatgaatggcttggagtgcactgagaacagcatatgatttcttgctttggggaaaaa 2177
Db 2112 ATGAATGAATGGCTTGGAGTGCACTGAGAACAGCATATGATTTCTTGCTTTGGGGAAAAA 2171
Qy 2178 gaatgatgctatgaaattggtgggtggtgtatttgagaagataatcattgcttatgtcaa 2237
Db 2172 GAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATAATCATTGCTTATGTCAA 2231
Qy 2238 atggagctgaatttgataaaaacccaaaatacagctatgaagtgctgggcaagtttactt 2297
Db 2232 ATGGAGCTGAATTTGATAAAAACCCAAAATACAGCTATGAAGTGCTGGGCAAGTTTACTT 2291
Qy 2298 tttttctgatgtttcctacaact 2320

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 21, 2002, 11:59:18 ; Search time 148.28 Seconds
(without alignments)
3843.200 Million cell updates/sec

Title: US-09-835-082-3
Perfect score: 2320
Sequence: 1 ctcagctgcagttctctgat.....ttctgatgtttcctacaact 2320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1633.4	70.4	2339	5	PCT-US92-00282-2	Sequence 2, Appli
	2	1609.2	69.4	2336	5	PCT-US92-00282-1	Sequence 1, Appli
	3	967.2	41.7	1561	5	PCT-US92-00282-25	Sequence 25, Appl
	4	344.4	14.8	2107	4	US-09-180-852-1	Sequence 1, Appli
	5	222.4	9.6	783	5	PCT-US92-00282-22	Sequence 22, Appl
	6	192.6	8.3	1008	5	PCT-US92-00282-10	Sequence 10, Appl
	7	186.6	8.0	1448	5	PCT-US92-00282-16	Sequence 16, Appl
	8	183.6	7.9	1066	5	PCT-US92-00282-14	Sequence 14, Appl
	9	161.8	7.0	1190	5	PCT-US92-00282-18	Sequence 18, Appl
	10	161.2	6.9	1197	5	PCT-US92-00282-12	Sequence 12, Appl
	11	143.6	6.2	1219	5	PCT-US92-00282-8	Sequence 8, Appli
	12	135	5.8	513	5	PCT-US92-00282-20	Sequence 20, Appl
	13	54.6	2.4	1669	2	US-08-522-421-1	Sequence 1, Appli
	14	44.8	1.9	1607	2	US-08-797-226-1	Sequence 1, Appli
	15	43.6	1.9	51	5	PCT-US92-00282-37	Sequence 37, Appl
	16	39.8	1.7	7218	1	US-08-232-463-14	Sequence 14, Appl
	17	38.2	1.6	1627	3	US-09-106-464-1	Sequence 1, Appli
	18	36.6	1.6	1731	2	US-08-466-583-1	Sequence 1, Appli
	19	36.6	1.6	1731	5	PCT-US95-07820-1	Sequence 1, Appli
c	20	36.6	1.6	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	21	36.6	1.6	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	22	36	1.6	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	23	36	1.6	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	24	35.8	1.5	1193	4	US-09-372-422A-23	Sequence 23, Appl
c	25	35.8	1.5	1696	1	US-07-612-674-1	Sequence 1, Appli
c	26	35.2	1.5	792	3	US-08-950-720A-1	Sequence 1, Appli
	27	35.2	1.5	1636	6	5447867-2	Patent No. 5447867
c	28	35.2	1.5	4839	4	US-09-353-332-1	Sequence 1, Appli
c	29	35.2	1.5	10607	1	US-08-078-090-3	Sequence 3, Appli
	30	35	1.5	1800	6	5180581-1	Patent No. 5180581
	31	35	1.5	2793	1	US-08-281-916-5	Sequence 5, Appli
	32	35	1.5	2793	2	US-08-460-725-7	Sequence 7, Appli
	33	35	1.5	10815	4	US-09-004-838-21	Sequence 21, Appl
	34	35	1.5	13149	4	US-09-004-838-87	Sequence 87, Appl
	35	34.6	1.5	1330	2	US-08-933-750C-80	Sequence 80, Appl
	36	34.6	1.5	1330	3	US-09-234-613-80	Sequence 80, Appl
c	37	34.6	1.5	1534	1	US-08-480-784-13	Sequence 13, Appl
c	38	34.6	1.5	1534	1	US-08-483-553-13	Sequence 13, Appl
c	39	34.6	1.5	1534	1	US-08-487-002-13	Sequence 13, Appl
c	40	34.6	1.5	1534	1	US-08-483-554B-13	Sequence 13, Appl
c	41	34.6	1.5	1534	1	US-08-488-011B-13	Sequence 13, Appl
c	42	34.6	1.5	1534	4	US-08-850-727-13	Sequence 13, Appl
c	43	34.6	1.5	1534	5	PCT-US95-10202-13	Sequence 13, Appl
c	44	34.6	1.5	1534	5	PCT-US95-10203-13	Sequence 13, Appl
c	45	34.6	1.5	1534	5	PCT-US95-10220-13	Sequence 13, Appl

ALIGNMENTS

```
RESULT   1
PCT-US92-00282-2
; Sequence 2, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2339 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-2


  Query Match             70.4%;  Score 1633.4;  DB 5;  Length 2339;
  Best Local Similarity   83.3%;  Pred. No. 0;
  Matches 1869;  Conservative    0;  Mismatches  371;  Indels    3;  Gaps    1;

Qy      81 ctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcaccat 140
           ||  || |||   || ||| |  |||| |||||||  ||||| |||| |||| ||| |||
Db      72 CTGGGCTGAGAGTGGAAAGGTGTTGGTGGTGCCCACTGATGGCAGCCCCTGGCTCAGCAT 131
```

```
Qy   1252 gttctggaaatgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaa 1311
          || |  ||||||||| |||| ||||| |||||||| || ||| | ||||||||  ||||
Db    421 gtccttgaaatgactgctgatgatttggaaaatgcccttaaaactgtcatcaacaacaag 480

Qy   1312 agttacaaggagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccg 1371
          || ||||||||||||||||||||||||||||||||||||||||||||||||||  | |||||
Db    481 agctacaaggagaacatcatgcgcctctccagccttcacaaggaccgtcctatagagcct 540

Qy   1372 ctggacctggccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctg 1431
          ||||||||||| |||||||||||||| |  |||||||||||||||||| || |||||||||
Db    541 ctggacctggctgtgttctgggtggaatacgtgatgaggcacaaggggcaccacacctg 600

Qy   1432 cgccccgcagcccacgacctcacctggtaccagtaccattccttggacgtgattggtttc 1491
          ||||| || ||||| |||||||||||||| |||||||| |||||||| |||||||| |||
Db    601 cgcccggccgcccatgacctcacctggtatcagtaccactccttggatgtgattggcttc 660

Qy   1492 ctcttggccgtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctac 1551
          ||| ||||| | ||| ||||||||| ||||||   |||||||||||||| ||||||| |
Db    661 ctcctggccattgtgttgacagtggtcttcattgnctttaaatgttgtgcctatggctgc 720

Qy   1552 cggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattga 1611
          |||||||||| |||  ||| |||| ||| |||||| | ||||||||||||||||||||
Db    721 cggaaatgctttgggggaaaggggcnagtgaagaaatcacacaaatccaagacccattga 780

Qy   1612 gaagtgggtgggaaataaggtaaaattttgaaccattccct 1652
          |||||||| || |   || |  || | ||    ||||  ||
Db    781 gaagtgggggaagtgaangagaagtattagttcattatct 821
```

Search completed: August 21, 2002, 12:46:07
Job time: 22285 sec

```
CC   animals or modified cells e.g. for pharmacogenetic screening.
XX
SQ   Sequence 759 BP; 190 A; 165 C; 169 G; 235 T; 0 other;


  Query Match             28.4%;  Score 659.8;  DB 21;  Length 759;
  Best Local Similarity   91.8%;  Pred. No. 4.4e-177;
  Matches  697;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy    115 atggatgggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggg 174
          |||||||||||| ||||||||||||||||  ||||||||||||||||||| || |||||||||
Db      1 atggatgggagtcactggttcaccatgcagtcggtggtggagaaacttatcctcaggggg 60

Qy    175 catgaggtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgc 234
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     61 catgaggtggttgtagtcatgccagaggtgagttggcaactggaaagatcactgaattgc 120

Qy    235 acagtgaagacttattcaacttcatataccctggaggatctggaccgggagttcaaggct 294
          |||||||||||||| ||||| || || || ||||| |||| | ||||||| |||| || |
Db    121 acagtgaagacttactcaacctcgtacactctggaagatcagaaccgggaattcatggtt 180

Qy    295 tttgcccatgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttca 354
          || ||||||||||||||||||||||| | || |||||||| ||||||||||||| |||||
Db    181 ttcgcccatgctcaatggaaagcacaggcacaaagtatatttttctctattaatgagttca 240

Qy    355 tacaatgacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaa 414
          | || ||   || ||||||||||||||||  ||||||||||||||||||| |||  ||||
Db    241 tccagtggttttcttgacttatttttttcgcattgcaggagtttgtttaatgaccgaaaa 300

Qy    415 ttagtagaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataac 474
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||| |
Db    301 ttagtagaatacttaaaggagagttcttttgatgcagtgtttctggatccttttgatacc 360

Qy    475 tgtggcttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaata 534
          |||||||||||||||| |||||||||||||||||||| ||||||||| ||||||||||
Db    361 tgtggcttaattgttgctaaatatttctccctcccctctgtggtcttcaccaggggaata 420

Qy    535 ctttgccactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccaga 594
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ttttgccaccatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccaat 480

Qy    595 attctcttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatg 654
            |||||||||||||||||||||||||||||||||||||||||||| |||||||||||| ||
Db    481 gatctcttagggttctcagatgccatgactttcaaggagagagtatggaaccacatcgtg 540

Qy    655 cacttggaggaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctct 714
          |||||||||| |||||||| |||||   | |||| | ||||||||||||||||||||||||
Db    541 cacttggaggaccatttattttgccagtatctttttagaaatgccctagaaatagcctct 600

Qy    715 gaaattctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttg 774
          |||||||||||||| ||||| ||||  |||||||||||||| ||||||||||||||||||
Db    601 gaaattctccaaacccctgtcacggcatatgatctctacagtcacacatcaatttggttg 660

Qy    775 ttgcgaacggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcatt 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ttgcgaacggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcatt 720

Qy    835 ggtggtatcaactgccatcagggaaagccgttgcctatg 873
          |||||||||||||| |||||||||||| |||||||||
Db    721 ggtggtatcaactgtcatcagggaaagccattgcctatg 759


RESULT  15
ABL01468
ID   ABL01468 standard; DNA; 874 BP.
XX
AC   ABL01468;
XX
DT   15-MAR-2002  (first entry)
XX
DE   Murine apoptosis related DNA sequence #133.
XX
KW   Apoptosis; mouse; cancer; autoimmune disease; viral infection;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW   reperfusion injury; stroke; liver damage; dilatory cardiomyopathy;
KW   transgenic animal; hepatotropic; antialcoholism; cytostatic;
KW   immunosuppressive; virucide; nootropic; neuroprotective; vasotropic;
KW   antiparkinsonian; cerebroprotective; ds.
XX
OS   Mus sp.
XX
PN   DE10126344-A1.
XX
PD   24-JAN-2002.
XX
PF   30-MAY-2001; 2001DE-1026344.
XX
PR   14-JUL-2000; 2000DE-1034303.
XX
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Grimm S,  Schoenfeld N,  Braziulis E,  Cramer U,  Gewies A,  Voss F;
PI   Mund T,  Albayrak T,  Gille H,  Klein M;
XX
DR   WPI; 2002-115563/16.
XX
PT   New apoptosis-associated nucleic acid sequences and polypeptides,
PT   useful for diagnosis, treatment and prevention of e.g. tumors and
PT   neurodegeneration -
XX
PS   Claim 1; Page 184; 227pp; German.
XX
CC   The present invention relates to nucleic acids from the mouse, where the
CC   nucleic acid is associated with apoptosis. The sequences can be used in
CC   the diagnosis, treatment and prevention of diseases associated with
CC   excessive or inadequate apoptosis, including tumours, autoimmune
CC   diseases, viral infections, degradative diseases (Alzheimer's,
CC   Parkinson's and Huntington's diseases), reperfusion injury, stroke and
CC   alcohol-induced injury to the liver, for identifying agents for treating
CC   these diseases, and to prepare transgenic animals in which expression of
CC   an apoptosis related sequence is altered. These are useful for genetic
CC   and/or pharmacological investigations of apoptosis and related diseases,
CC   including dilatory cardiomyopathy. The present sequence is one of the
CC   apoptosis related sequences of the invention.
XX
SQ   Sequence 874 BP; 222 A; 202 C; 233 G; 210 T; 7 other;


  Query Match             24.9%;  Score 578.8;  DB 24;  Length 874;
  Best Local Similarity   84.9%;  Pred. No. 5.3e-154;
  Matches  646;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy    892 aatgcttctggagaacatggaattgtggttttctctttgggatcaatggtctcagaaatt 951
          || || ||||| || ||||| || |||||||||||||||||||| |||||||||| |||
Db     61 aacgcctctggggagcatggcatcgtggttttctctttgggatccatggtctcagagatt 120

Qy    952 ccagagaagaaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctg 1011
          || |||||||||| |||| ||||||||| |||||||||| ||| |||||||| ||||||
Db    121 ccggagaagaaagccatggaaattgctgaggctttgggcagaattcctcagacggtcctg 180

Qy   1012 tggcggtacactggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtgg 1071
          ||||| |||| |||||  ||||||||||||||| || |||| || ||||| || |||
Db    181 tggcgctacaccggaactagaccatcgaatcttgcaaagaacacaattcttgtcaaatgg 240

Qy   1072 ctaccccaaaacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggt 1131
          |||||||||| |||||||||||||| || | ||| || || || ||||| ||  ||||
Db    241 ctaccccaaaatgatctgcttggtcatccaaagactcgggcattcatcacacactctggc 300

Qy   1132 tcccatggtgtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgttt 1191
          ||||||||| |||||||| | |||||||||| ||||| ||||||||||||||| | |||
Db    301 tcccatggtatttatgaaggaatatgcaatggagttccgatggtgatgatgcccctattt 360

Qy   1192 ggtgatcagatggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaat 1251
          || |||||||||||||||| |||||||||| |||  ||||||||| ||||||||||||
Db    361 ggcgatcagatggacaatgccaagcgcatggaaactcggggagctggggtgaccctgaat 420
```

```
XX
DR   WPI; 2000-052981/04.
DR   P-PSDB; AAY57100.
XX
PT   New nucleic acid representing polymorphisms in the human uridine
PT   diphosphate glucuronosyltransferase gene, used for diagnosis and
PT   evaluation of drug metabolism  -
XX
PS   Examples; Page 43-44; 63pp; English.
XX
CC   Nucleotide sequences AAZ45110-Z45118 are exons 1A-1J of the human
CC   uridine diphosphate-glucuronosyltransferase 1 (UGT1) gene. The UGTs are a
CC   family of enzymes that catalyse the glucuronic acid conjugation of a wide
CC   range of endogenous and exogenous substrates including  phenols,
CC   alcohols, amines and fatty acids. Many of the reactions catalysed by UGTs
CC   result in toxic substances being converted to compounds which are more
CC   water soluble and are excreted. The invention relates to and identifies
CC   UGT1 polymorphisms (AAZ45004-Z45041). The polymorphism sequences are
CC   useful as probes for detecting UGT1 locus polymorphisms, indicative of
CC   altered UGT1 expression or activity. These polymorphisms are associated
CC   with Crigler-Najjar and Gilbert syndromes (unconjugated
CC   hyperbilirubinaemia) and drug metabolism. The genotyping of the UGT1 gene
CC   is used to predict the rate of metabolism of UGT1 substrates, possible
CC   drug-drug interactions and adverse side effects (i.e. to optimize drug
CC   dosage), and to screen for diseases caused by exposure to toxins and to
CC   study the effects of polymorphisms on enzymatic activity. The UGT1
CC   sequences, including polymorphisms, can also be used to produce the
CC   corresponding protein (or its fragments) or to generate transgenic
CC   animals or modified cells e.g. for pharmacogenetic screening.
XX
SQ   Sequence 735 BP; 189 A; 179 C; 187 G; 180 T; 0 other;


  Query Match             31.7%;  Score 735;  DB 21;  Length 735;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-198;
  Matches  735;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    874 gaatttgaagcctacattaatgcttctggagaacatggaattgtggttttctctttggga 933
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 gaatttgaagcctacattaatgcttctggagaacatggaattgtggttttctctttggga 60

Qy    934 tcaatggtctcagaaattccagagaagaaagctatggcaattgctgatgctttgggcaaa 993
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 tcaatggtctcagaaattccagagaagaaagctatggcaattgctgatgctttgggcaaa 120

Qy    994 atccctcagacagtcctgtggcggtacactggaacccgaccatcgaatcttgcgaacaac 1053
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 atccctcagacagtcctgtggcggtacactggaacccgaccatcgaatcttgcgaacaac 180

Qy   1054 acgatacttgttaagtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcc 1113
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 acgatacttgttaagtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcc 240

Qy   1114 tttatcacccatgctggttcccatggtgtttatgaaagcatatgcaatggcgttcccatg 1173
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 tttatcacccatgctggttcccatggtgtttatgaaagcatatgcaatggcgttcccatg 300

Qy   1174 gtgatgatgcccttgtttggtgatcagatggacaatgcaaagcgcatggagactaaggga 1233
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 gtgatgatgcccttgtttggtgatcagatggacaatgcaaagcgcatggagactaaggga 360

Qy   1234 gctgcagtgaccctgaatgttctggaaatgacttctgaagatttagaaaatgctctaaaa 1293
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gctgcagtgaccctgaatgttctggaaatgacttctgaagatttagaaaatgctctaaaa 420

Qy   1294 gcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctccagccttcacaag 1353
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 gcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctccagccttcacaag 480

Qy   1354 gacccccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgaggcac 1413
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 gacccccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgaggcac 540

Qy   1414 aagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccattcc 1473
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 aagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccattcc 600

Qy   1474 ttggacgtgattggtttcctcttggccgtcgtgctgacagtggccttcatcacctttaaa 1533
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ttggacgtgattggtttcctcttggccgtcgtgctgacagtggccttcatcacctttaaa 660

Qy   1534 tgttgtgctcatggctaccggaaatgcttggggaaaaaagggcgagttaagaaagcccac 1593
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 tgttgtgctcatggctaccggaaatgcttggggaaaaaagggcgagttaagaaagcccac 720

Qy   1594 aaatccaagacccat 1608
          ||||||||||||||
Db    721 aaatccaagacccat 735


RESULT  14
AAZ45117
ID   AAZ45117 standard; DNA; 759 BP.
XX
AC   AAZ45117;
XX
DT   28-FEB-2000  (first entry)
XX
DE   UDP-glucuronosyltransferase 1 (UGT1) exon 1J nucleotide sequence.
XX
KW   Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
KW   glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
KW   unconjugated hyperbilirubinaemia; drug metabolism; transgenic animal;
KW   pharmacogenetic screening; diagnose; ss.
XX
OS   Homo sapiens.
XX
PN   WO9957322-A2.
XX
PD   11-NOV-1999.
XX
PF   04-MAY-1999;   99WO-US09702.
XX
PR   07-MAY-1998;   98US-0084807.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Penny L,  Galvin M;
XX
DR   WPI; 2000-052981/04.
DR   P-PSDB; AAY57099.
XX
PT   New nucleic acid representing polymorphisms in the human uridine
PT   diphosphate glucuronosyltransferase gene, used for diagnosis and
PT   evaluation of drug metabolism  -
XX
PS   Examples; Page 42; 63pp; English.
XX
CC   Nucleotide sequences AAZ45110-Z45118 are exons 1A-1J of the human
CC   uridine diphosphate-glucuronosyltransferase 1 (UGT1) gene. The UGTs are a
CC   family of enzymes that catalyse the glucuronic acid conjugation of a wide
CC   range of endogenous and exogenous substrates including  phenols,
CC   alcohols, amines and fatty acids. Many of the reactions catalysed by UGTs
CC   result in toxic substances being converted to compounds which are more
CC   water soluble and are excreted. The invention relates to and identifies
CC   UGT1 polymorphisms (AAZ45004-Z45041). The polymorphism sequences are
CC   useful as probes for detecting UGT1 locus polymorphisms, indicative of
CC   altered UGT1 expression or activity. These polymorphisms are associated
CC   with Crigler-Najjar and Gilbert syndromes (unconjugated
CC   hyperbilirubinaemia) and drug metabolism. The genotyping of the UGT1 gene
CC   is used to predict the rate of metabolism of UGT1 substrates, possible
CC   drug-drug interactions and adverse side effects (i.e. to optimize drug
CC   dosage), and to screen for diseases caused by exposure to toxins and to
CC   study the effects of polymorphisms on enzymatic activity. The UGT1
CC   sequences, including polymorphisms, can also be used to produce the
CC   corresponding protein (or its fragments) or to generate transgenic
```

```
DE   UDP-glucuronosyltransferase 1 (UGT1) exon 1G nucleotide sequence.
XX
KW   Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
KW   glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
KW   unconjugated hyperbilirubinaemia; drug metabolism; transgenic anima1;
KW   pharmacogenetic screening; diagnose; ss.
XX
OS   Homo sapiens.
XX
PN   WO9957322-A2.
XX
PD   11-NOV-1999.
XX
PF   04-MAY-1999;   99WO-US09702.
XX
PR   07-MAY-1998;   98US-0084807.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Penny L,  Galvin M;
XX
DR   WPI; 2000-052981/04.
DR   P-PSDB; AAY57097.
XX
PT   New nucleic acid representing polymorphisms in the human uridine
PT   diphosphate glucuronosyltransferase gene, used for diagnosis and
PT   evaluation of drug metabolism -
XX
PS   Examples; Page 37-38; 63pp; English.
XX
CC   Nucleotide sequences AAZ45110-Z45118 are exons 1A-1J of the human
CC   uridine diphosphate-glucuronosyltransferase 1 (UGT1) gene. The UGTs are a
CC   family of enzymes that catalyse the glucuronic acid conjugation of a wide
CC   range of endogenous and exogenous substrates including  phenols,
CC   alcohols, amines and fatty acids. Many of the reactions catalysed by UGTs
CC   result in toxic substances being converted to compounds which are more
CC   water soluble and are excreted. The invention relates to and identifies
CC   UGT1 polymorphisms (AAZ45004-Z45041). The polymorphism sequences are
CC   useful as probes for detecting UGT1 locus polymorphisms, indicative of
CC   altered UGT1 expression or activity. These polymorphisms are associated
CC   with Crigler-Najjar and Gilbert syndromes (unconjugated
CC   hyperbilirubinaemia) and drug metabolism. The genotyping of the UGT1 gene
CC   is used to predict the rate of metabolism of UGT1 substrates, possible
CC   drug-drug interactions and adverse side effects (i.e. to optimize drug
CC   dosage), and to screen for diseases caused by exposure to toxins and to
CC   study the effects of polymorphisms on enzymatic activity. The UGT1
CC   sequences, including polymorphisms, can also be used to produce the
CC   corresponding protein (or its fragments) or to generate transgenic
CC   animals or modified cells e.g. for pharmacogenetic screening.
XX
SQ   Sequence 951 BP; 229 A; 211 C; 214 G; 297 T; 0 other;


  Query Match            31.9%;  Score 741.2;  DB 21;  Length 951;
  Best Local Similarity  90.5%;  Pred. No. 3.5e-200;
  Matches  791;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy     19 atggcttgcacagggtggaccagccccttcctctatgtgtgtctgctgctgacctgt 78
          ||||||  |  ||||||||||  ||| ||||||  |||| |||||||||  ||||||||||||
Db      1 atggctcgtgcagggtggactggcctccttcccctatatgtgtgtctactgctgacctgt 60

Qy     79 ggctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcacc 138
          |  ||  |   | ||||||||||  ||||||||||||||||||||||||||||||||||
Db     61 gctttgccaaggtcagggaagctgctggtagtgcccatggatgggagccactggttcacc 120

Qy    139 atgaggtcggtggtggagaaactcattctcaggggggcatgaggtggttgtagtcatgcca 198
          |||  ||||||||||||||||||||| ||||||||||||||||||| ||||||||||||
Db    121 atgcagtcggtggtggagaaactcatcctcaggggggcatgaggtggtcgtagtcatgcca 180

Qy    199 gaggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttca 258
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||| |||
Db    181 gaggtgagttggcaactgggaagatcactgaattgcacagtgaagacttactcaacctca 240

Qy    259 tataccctggaggatctggaccgggagttcaaggcttttgcccatgctcaatggaaagca 318
          || || |||||||||| |||||||||||||| || ||||||| ||||||  ||||  |||
Db    241 tacactctggaggatcaggaccgggagttcatggtttttgccgatgctcgctggacggca 300

Qy    319 caagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttattt 378
          | | | ||||||  || ||||||||||   |||||| |||||  ||||||||||||||||
Db    301 ccattgcgaagtgcatttttctctattaacaagttcatccaatggtatttttgacttattt 360

Qy    379 ttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagagt 438
          |||||||||||||||||||||||||| |||  ||||||||||||||||||||||||||||
Db    361 ttttcaaattgcaggagtttgtttaatgaccgaaaattagtagaatacttaaaggagagt 420

Qy    439 tcttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaatat 498
          | ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db    421 tgttttgatgcagtgtttctcgatccttttgatcgctgtggcttaattgttgccaaatat 480

Qy    499 ttctccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaaggt 558
          |||||||||||||| |||||||||||||||||||| ||||||||||||||||||||||||
Db    481 ttctccctccctctgtggtcttcgccaggggaatattttgccactatcttgaagaaggt 540

Qy    559 gcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgcc 618
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||| |||
Db    541 gcacagtgccctgctcctctttcctatgtcccagacttctcttagggttctcagacgcc 600

Qy    619 atgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatgc 678
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db    601 atgactttcaaggagagagtatggaaccacatcatgcacttggaggaacatttattttgc 660

Qy    679 caccgttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttacg 738
          | |  |||||||||||||| | ||||||||||||||||||||||||||| ||||| |||
Db    661 ccctatttttttcaaaaatgtcttagaaatagcctctgaaattctccaaacccctgtcacg 720

Qy    739 gagtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttggac 798
          |  ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    721 gcatatgatctctacagccacacatcaatttggttgttgcgaactgactttgttttggag 780

Qy    799 tatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggga 858
          |||||||||||||||||||||| ||||||||||||||||||||||| |||||||||
Db    781 tatcccaaacccgtgatgcccaatatgatcttcattggtggtatcaactgtcatcaggga 840

Qy    859 aagccgttgcctatggaatttgaagcctacatta 892
          |||||  |||||||||  |  | |   |  | |||
Db    841 aagccagtgcctatggtaagttatctccccttta 874


RESULT  13
AAZ45118
ID   AAZ45118 standard; DNA; 735 BP.
XX
AC   AAZ45118;
XX
DT   28-FEB-2000  (first entry)
XX
DE   UDP-glucuronosyltransferase 1 (UGT1) exons 2-5 nucleotide sequence.
XX
KW   Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
KW   glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
KW   unconjugated hyperbilirubinaemia; drug metabolism; transgenic animal;
KW   pharmacogenetic screening; diagnose; ss.
XX
OS   Homo sapiens.
XX
PN   WO9957322-A2.
XX
PD   11-NOV-1999.
XX
PF   04-MAY-1999;   99WO-US09702.
XX
PR   07-MAY-1998;   98US-0084807.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Penny L,  Galvin M;
```

```
Db    601 catatgatttcttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatt 660

Qy   2211 tgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatataca 2270
Db    661 tgagaagataatcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatataca 720

Qy   2271 gctatgaagtgctgggcaagtttacttttttttctgatgtttcctacaact 2320
Db    721 gctatgaagtgctgggcaagtttacttttttttctgatgtttcctacaact 770
```

RESULT 11
AAZ45116
ID AAZ45116 standard; DNA; 930 BP.
XX
AC AAZ45116;
XX
DT 28-FEB-2000 (first entry)
XX
DE UDP-glucuronosyltransferase 1 (UGT1) exon 1H nucleotide sequence.
XX
KW Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
KW glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
KW unconjugated hyperbilirubinaemia; drug metabolism; transgenic animal;
KW pharmacogenetic screening; diagnose; ss.
XX
OS Homo sapiens.
XX
PN WO9957322-A2.
XX
PD 11-NOV-1999.
XX
PF 04-MAY-1999; 99WO-US09702.
XX
PR 07-MAY-1998; 98US-0084807.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Penny L, Galvin M;
XX
DR WPI; 2000-052981/04.
DR P-PSDB; AAY57098.
XX
PT New nucleic acid representing polymorphisms in the human uridine
PT diphosphate glucuronosyltransferase gene, used for diagnosis and
PT evaluation of drug metabolism -
XX
PS Examples; Page 39-41; 63pp; English.
XX
CC Nucleotide sequences AAZ45110-Z45118 are exons 1A-1J of the human
CC uridine diphosphate-glucuronosyltransferase 1 (UGT1) gene. The UGTs are a
CC family of enzymes that catalyse the glucuronic acid conjugation of a wide
CC range of endogenous and exogenous substrates including phenols,
CC alcohols, amines and fatty acids. Many of the reactions catalysed by UGTs
CC result in toxic substances being converted to compounds which are more
CC water soluble and are excreted. The invention relates to and identifies
CC UGT1 polymorphisms (AAZ45004-Z45041). The polymorphism sequences are
CC useful as probes for detecting UGT1 locus polymorphisms, indicative of
CC altered UGT1 expression or activity. These polymorphisms are associated
CC with Crigler-Najjar and Gilbert syndromes (unconjugated
CC hyperbilirubinaemia) and drug metabolism. The genotyping of the UGT1 gene
CC is used to predict the rate of metabolism of UGT1 substrates, possible
CC drug-drug interactions and adverse side effects (i.e. to optimize drug
CC dosage), and to screen for diseases caused by exposure to toxins and to
CC study the effects of polymorphisms on enzymatic activity. The UGT1
CC sequences, including polymorphisms, can also be used to produce the
CC corresponding protein (or its fragments) or to generate transgenic
CC animals or modified cells e.g. for pharmacogenetic screening.
XX
SQ Sequence 930 BP; 228 A; 205 C; 212 G; 285 T; 0 other;

Query Match 33.0%; Score 766.6; DB 21; Length 930;
Best Local Similarity 91.6%; Pred. No. 2.1e-207;
Matches 811; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```
Qy     19 atggcttgcacagggtggaccagccccttcctctatgtgtgtgtctgctgctgacctgt 78
Db      1 atggctcgcacagggtggaccagccccattcccctatgtgtttctctgctgctgacctgt 60

Qy     79 ggctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcacc 138
Db     61 ggctttgctgaggcagggaagctgctggtagtgcccatggatgggagtcactggttcacc 120

Qy    139 atgaggtcggtggtggagaaactcattctcaggggcatgaggtggttgtagtcatgcca 198
Db    121 atgcagtcggtggtggagaaacttatcctcaggggcatgaggtggttgtagtcatgcca 180

Qy    199 gaggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttca 258
Db    181 gaggtgagttggcaactgggaaaatcactgaattgcacagtgaagacttactcaacctca 240

Qy    259 tataccctggaggatctggaccgggagttcaaggcttttgcccatgctcaatggaaagca 318
Db    241 tacactctggaggatctggaccgggaattcatggatttcgccgatgctcaatggaaagca 300

Qy    319 caagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttattt 378
Db    301 caagtacgaagtttgttttctctatttctgagttcatccaatggttttttaacttattt 360

Qy    379 ttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagagt 438
Db    361 ttttcgcattgcaggagtttgtttaatgaccgaaaattagtagaatacttaaaggagagt 420

Qy    439 tcttttgatgcagtgtttctcgatcctttgataactgtggcttaattgttgccaaatat 498
Db    421 tcttttgatgcggtgtttcttgatcctttgatgcctgtgcgttaattgttgccaaatat 480

Qy    499 ttctccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaaggt 558
Db    481 ttctccctcccctctgtggtcttcgccaggggaataggttgccactatcttgaagaaggt 540

Qy    559 gcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgcc 618
Db    541 gcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgcc 600

Qy    619 atgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatgc 678
Db    601 atgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattttgc 660

Qy    679 caccgtttttttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttacg 738
Db    661 cagtatttttccaaaaatgccctagaaatagcctctgaaattctccaaacacctgtcaca 720

Qy    739 gagtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttggac 798
Db    721 gcatatgatctctacagccacacatcaatttggttgttgcgaacagactttgttttggac 780

Qy    799 tatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggga 858
Db    781 tatcccaaacccgtgatgcccaatatgatcttcattggtggtatcaactgccatcaggga 840

Qy    859 aagccgttgcctatggaatttgaagcctacattaatgcttctgga 903
Db    841 aagccattgcctatggtaagtcacctctcctttagcacattagga 885
```

RESULT 12
AAZ45115
ID AAZ45115 standard; DNA; 951 BP.
XX
AC AAZ45115;
XX
DT 28-FEB-2000 (first entry)
XX

```
Qy   2010 aattgatccttaaccaataatggtcagtcctcatctctgtcctgcttcataggtgccacc 2069
          ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    980 aattgatccttaaccaataatggtcagtcctcatctctgtcgtgcttcataggtgccacc 1039

Qy   2070 ttgtgtgtttaaagaagggaagctttgtacctttagagtgtag---------------- 2112
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1040 ttgtgtgtttaaagaagggaagctttgtacctttagagtgtagtgaaatgaatgaatggc 1099

Qy   2113 ------------------------------------------------------------ 2112

Db   1100 ttggagtgcactgagaacagcatatgatttcttgctttggggaaaaagaatgatgctatg 1159

Qy   2113 ------------------gtgaaatgaatgaatggcttggagtgcactgagaacagcata 2154
                            ||||||||||||||||||||||||||||||||||||||||||
Db   1160 aaattggtgggtggtgtagtgaaatgaatgaatggcttggagtgcactgagaacagcata 1219

Qy   2155 tgatttcttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgag 2214
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1220 tgatttcttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgag 1279

Qy   2215 aagataatcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagcta 2274
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1280 aagataatcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagcta 1339

Qy   2275 tgaagtgctgggcaagtttacttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||||||||||
Db   1340 tgaagtgctgggcaagtttacttttttctgatgtttcctacaact 1385
```

RESULT 10
AAH51405
ID AAH51405 standard; DNA; 1001 BP.
XX
AC AAH51405;
XX
DT 29-AUG-2001 (first entry)
XX
DE Human UGT1A7 related DNA containing a biallelic polymorphism SEQ ID 296.
XX
KW Human; biallelic marker; single nucleotide polymorphism; SNP; MGSTII;
KW microsomal glutathione S-transferase II; malate decarboxylase enzyme;
KW DME1; ME1; cytochrome P450; glutathione reductase; GSHR; GSHS; GGT5;
KW flavin-containing monooxygenase; FMO; gamma-glutamyltransferase 5;
KW dipeptidase; DP; glucose 6-phosphate dehydrogenase; G6PDH; haplotype;
KW phosphogluconate dehydrogenase; PGDH; drug metabolism; phenotype;
KW uridine diphosphate glucoronosyl transferase; UGT2; asthma; hepatoxicity;
KW zileuton; ds.
XX
OS Homo sapiens.
XX
PN WO200058508-A2.
XX
PD 05-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-IB00403.
XX
PR 25-MAR-1999; 99US-0126269.
PR 30-APR-1999; 99US-0131961.
XX
PA (GEST) GENSET.
XX
PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A;
XX
DR WPI; 2000-638353/61.
XX
PT Polynucleotides comprising sequences from malate decarboxylase
PT enzyme-related biallelic markers used for genotyping -
XX
PS Claim 13; Page 488-489; 673pp; English.
XX
CC Sequences AAH51110-AAH51593 represent human DNA fragments which contain
CC biallelic markers. The sequences are related to various human genes
CC including microsomal glutathione S-transferase II (MGSTII), malate
CC decarboxylase enzyme (DME1/ME1), cytochrome P450, glutathione
CC reductase/synthase (GSHR/GSHS), flavin-containing monooxygenases (FMO),
CC gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 6-phosphate
CC dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine
CC diphosphate glucoronosyl transferases (UGT2). Each of these sequences
CC contains a biallelic marker/polymorphism, which is represented in the
CC sequence as a degenerate/undefined base. The genes to which the biallelic
CC marker containing sequences are related are involved in drug metabolism.
CC Sequences AAH51594 - AAH51598 represent the genomic sequence of the
CC MGSTII gene and four alternative MGSTII cDNA sequences. AAB62905-AAB62906
CC are MGSTII gene products. PCR primers AAH51599 and AAH51600 are used in
CC an example for the amplification of human genomic DNA fragments. The
CC invention includes a method of genotyping comprising determining the
CC identity of a nucleotide at a DME- or MGSTII-related biallelic marker in
CC a biological sample. The method is used to determine the frequency in
CC population of an allele of a DME- or MGST-II related biallelic marker and
CC to select an individual for inclusion in a clinical trial of a drug
CC treatment. The method is also used to detect association between allele
CC and phenotype, and to detect association between haplotype and phenotype.
CC The polynucleotides are used, in hybridization assays, sequencing assays
CC or allele specific amplification assays. The method can be used to
CC determine whether an individual suffers or is at risk of developing
CC asthma or is at risk of developing hepatoxicity on treatment with
CC zileuton.
XX
SQ Sequence 1001 BP; 315 A; 169 C; 205 G; 311 T; 1 other;

Query Match 33.2%; Score 769.6; DB 21; Length 1001;
Best Local Similarity 99.9%; Pred. No. 3e-208;
Matches 769; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1551 ccggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattg 1610
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ccggaaatgcttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattg 60

Qy   1611 agaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttccaaacttga 1670
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 agaagtgggtgggaaataaggtaaaattttgaaccattccctagtcatttccaaacttga 120

Qy   1671 aaacagaatcagtgttaaattcattttattcttattaaggaaatactttgcataaattaa 1730
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 aaacagaatcagtgttaaattcattttattcttattaaggaaatactttgcataaattaa 180

Qy   1731 tcagccccagagtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagt 1790
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 tcagccccagagtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagt 240

Qy   1791 aaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgcc 1850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 aaagatatttgaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgcc 300

Qy   1851 atttttcagaggacgtgcagacaggctggcattctagattacttttcttactctgaaaca 1910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 atttttcagaggacgtgcagacaggctggcattctagattacttttcttactctgaaaca 360

Qy   1911 tggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatgg 1970
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 tggcctgtttgggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatgg 420

Qy   1971 cagttttaatcttatctttggcttctgcagatggttgcaattgatccttaaccaataat 2030
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 cagttttaatcttatctttggcttctgcagatggttgcaattgatccttaaccaataat 480

Qy   2031 ggtcagtcctcatctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaa 2090
          ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db    481 ggtcagtcctcatctctgtcstgcttcataggtgccaccttgtgtgtttaaagaagggaa 540

Qy   2091 gctttgtacctttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacag 2150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 gctttgtacctttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacag 600

Qy   2151 catatgatttcttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatt 2210
```

```
Db    541 tggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatctctgt 600

Qy   2050 cctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacctttagagtg 2109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 cctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacctttagagtg 660

Qy   2110 taggcgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttcttgctttg 2169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 taggcgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttcttgctttg 720

Qy   2170 gggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatcattgct 2229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 gggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatcattgct 780

Qy   2230 tatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgctgggcaa 2289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 tatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtgctgggcaa 840

Qy   2290 gtttacttttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||
Db    841 gtttacttttttttctgatgtttcctacaact 871


RESULT   9
AAQ33027
ID   AAQ33027 standard; DNA; 1667 BP.
XX
AC   AAQ33027;
XX
DT   27-JAN-1993  (first entry)
XX
DE   Exon 5 from the UGT1 gene locus.
XX
KW   UGT1A; UGT1BP; UGT1C; UGT1D; UGT1E; UGT1F; isozyme; bilirubin;
KW   UDP-glucuronosyl transferase; CN; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        1..20
FT                   /*tag=  a
FT                   /note= "representation of 3.3 kbp of
FT                          non-sequenced DNA between the sequences
FT                          represented in AAQ33026 and AAQ33027"
FT   exon            291..1402
FT                   /*tag=  b
FT                   /label= Exon_5
FT   misc_RNA        291..586
FT                   /*tag=  c
FT                   /note= "encodes part of transferase C-terminal; see CC"
FT   polyA_signal    1389..1394
FT                   /*tag=  d
XX
PN   WO9212987-A.
XX
PD   06-AUG-1992.
XX
PF   10-JAN-1992;   92WO-US00282.
XX
PR   10-JAN-1991;   91US-0639453.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Owens IS,  Ritter JK;
XX
DR   WPI; 1992-284593/34.
DR   P-PSDB; AAR30166.
XX
PT   Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT   for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT   types I and II
XX
PS   Disclosure; Fig 1I; 99pp; English.
```

```
XX
CC   The isolated gene locus, UGT1, has a sequence of about 10000 bp
CC   which represent (1) Exon 1, comprising 6 transcriptional units
CC   (UGT1F, E, D, C, BP and A), represented in AAQ27368 and
CC   AAQ33020-24 respectively;
CC                  (2) Exon 2, represented in AAQ33025;
CC                  (3) Exon 3, represented in AAQ33026;
CC                  (4) Exon 4, represented in AAQ33026;
CC                  (5) Exon 5, represented in AAQ33027; and
CC                  (6) about 69 kb of non-sequenced DNA.
CC   Six unique N-termini of 286-289 amino acids are encoded by
CC   the six different first exons and identical C-termini of 246 amino
CC   acids are encoded by the common exons 2-5.  The UGT1 gene locus
CC   encodes a family of UDP-glucuronosyl transferase isozymes, two of
CC   which metabolise bilirubin.
CC   Patients having Crigler-Najjar Syndrome (CN) Type I, have a
CC   mutation present in the second common exon.
XX
SQ   Sequence 1667 BP; 487 A; 316 C; 370 G; 474 T; 20 other;


  Query Match             34.4%;  Score 798;  DB 13;  Length 1667;
  Best Local Similarity   89.1%;  Pred. No. 3.4e-216;
  Matches 1003;  Conservative    0;  Mismatches   20;  Indels  103;  Gaps    9;

Qy   1291 aaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctccagccttcac 1350
          || | |    ||  ||     ||||||||||||||||||||||||||||||||||||||||
Db    267 aacgaaactgtctttgtgtttagttacaaggagaacatcatgcgcctctccagccttcac 326

Qy   1351 aaggaccgcccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgagg 1410
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    327 aaggaccgcccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgagg 386

Qy   1411 cacaagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccat 1470
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    387 cacaagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccat 446

Qy   1471 tccttggacgtgattggtttcctcttggccgtcgtgct-gacagtggccttcatcacctt 1529
          ||||||||||||||||||||||||||||||||||  ||| ||||||||||||||||||||
Db    447 tccttggacgtgattggtttcctcttggccgtctggctggacagtggccttcatcacctt 506

Qy   1530 taaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcgagttaagaaagc 1589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    507 taaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcgagttaagaaagc 566

Qy   1590 ccacaaatccaagacccattgagaagtgggtgggaaataaggtaaaattttgaaccattc 1649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    567 ccacaaatccaagacccattgagaagtgggtgggaaataaggtaaaattttgaaccattc 626

Qy   1650 cctagtcatttccaaacttgaaaacagaatcagtgttaaattcattttattcttattaag 1709
          |||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db    627 cctagtcatttcc-aacttgaaaacagaatcagtgttaaattcattttattcttattaag 685

Qy   1710 gaaatactttgcataaattaatcagccccagagtgctttaaaaaattctcttaaataaaa 1769
          |||||||  |||||||||||||||||||  ||||||||||||||||||||||||||||||
Db    686 gaaatac-ttgcataaattaatcagcccc-gagtgctttaaaaaattctcttaaataaaa 743

Qy   1770 ataatagactcgctagtcagtaaagatatttgaatatgtatcgtgccccctccggtgtct 1829
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||| |||
Db    744 ataatagactcgctagtcagtaaagatatttgaatatgtatcgtgccccctctggtctct 803

Qy   1830 ttgatcaggatgacatgtgccatttttcagaggacgtgcagacaggctggcattctagat 1889
          ||||||||  ||||||||||||||||||||||||||||||||||||| |||||||||||
Db    804 ttgatcagcatgacatgtgccatttttcagaggacgtgcagacaggctgg-attctagat 862

Qy   1890 tacttttcttactctgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccac 1949
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    863 tacttttcttactctgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccac 922

Qy   1950 cgctgcccctactgcaaatggcagttttaatcttatcttttggcttctgcagatggttgc 2009
           ||| |||||||||||||||| ||||||||||||||||||||||| |||||||
Db    923 ggct-cccctactgcaaatggcag-tttaatcttatcttttggcttctgcag-tggttgc 979
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 gacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttg 600

Qy   1922 ggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatc 1981
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatc 660

Qy   1982 ttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctc 2041
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctc 720

Qy   2042 atctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacct 2101
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 atctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacct 780

Qy   2102 ttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttc 2161
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttc 840

Qy   2162 ttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataa 2221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataa 900

Qy   2222 tcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtg 2281
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 tcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagtg 960

Qy   2282 ctgggcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||||||
Db    961 ctgggcaagtttacttttttttctgatgtttcctacaact 999
```

RESULT 8
AAH51404
ID AAH51404 standard; DNA; 1001 BP.
XX
AC AAH51404;
XX
DT 29-AUG-2001 (first entry)
XX
DE Human UGT1A7 related DNA containing a biallelic polymorphism SEQ ID 295.
XX
KW Human; biallelic marker; single nucleotide polymorphism; SNP; MGSTII;
KW microsomal glutathione S-transferase II; malate decarboxylase enzyme;
KW DME1; ME1; cytochrome P450; glutathione reductase; GSHR; GSHS; GGT5;
KW flavin-containing monooxygenase; FMO; gamma-glutamyltransferase 5;
KW dipeptidase; DP; glucose 6-phosphate dehydrogenase; G6PDH; haplotype;
KW phosphogluconate dehydrogenase; PGDH; drug metabolism; phenotype;
KW uridine diphosphate glucoronosyl transferase; UGT2; asthma; hepatoxicity;
KW zileuton; ds.
XX
OS Homo sapiens.
XX
PN WO200058508-A2.
XX
PD 05-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-IB00403.
XX
PR 25-MAR-1999; 99US-0126269.
PR 30-APR-1999; 99US-0131961.
XX
PA (GEST) GENSET.
XX
PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A;
XX
DR WPI; 2000-638353/61.
XX
PT Polynucleotides comprising sequences from malate decarboxylase
PT enzyme-related biallelic markers used for genotyping -
XX
PS Claim 13; Page 487-488; 673pp; English.
XX

CC Sequences AAH51110-AAH51593 represent human DNA fragments which contain
CC biallelic markers. The sequences are related to various human genes
CC including microsomal glutathione S-transferase II (MGSTII), malate
CC decarboxylase enzyme (DME1/ME1), cytochrome P450, glutathione
CC reductase/synthase (GSHR/GSHS), flavin-containing monooxygenases (FMO),
CC gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 6-phosphate
CC dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine
CC diphosphate glucoronosyl transferases (UGT2). Each of these sequences
CC contains a biallelic marker/polymorphism, which is represented in the
CC sequence as a degenerate/undefined base. The genes to which the biallelic
CC marker containing sequences are related are involved in drug metabolism.
CC Sequences AAH51594 - AAH51598 represent the genomic sequence of the
CC MGSTII gene and four alternative MGSTII cDNA sequences. AAB62905-AAB62906
CC are MGSTII gene products. PCR primers AAH51599 and AAH51600 are used in
CC an example for the amplification of human genomic DNA fragments. The
CC invention includes a method of genotyping comprising determining the
CC identity of a nucleotide at a DME- or MGSTII-related biallelic marker in
CC a biological sample. The method is used to determine the frequency in
CC population of an allele of a DME- or MGST-II related biallelic marker and
CC to select an individual for inclusion in a clinical trial of a drug
CC treatment. The method is also used to detect association between allele
CC and phenotype, and to detect association between haplotype and phenotype.
CC The polynucleotides are used, in hybridization assays, sequencing assays
CC or allele specific amplification assays. The method can be used to
CC determine whether an individual suffers or is at risk of developing
CC asthma or is at risk of developing hepatoxicity on treatment with
CC zileuton.
XX
SQ Sequence 1001 BP; 293 A; 171 C; 210 G; 326 T; 1 other;

Query Match 37.5%; Score 870.6; DB 21; Length 1001;
Best Local Similarity 99.9%; Pred. No. 5.6e-237;
Matches 870; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1450 ctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcgtgctg 1509
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcgtgctg 60

Qy   1510 acagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaa 1569
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 acagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaa 120

Qy   1570 aaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtgggaaataa 1629
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 aaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtgggaaataa 180

Qy   1630 ggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgttaaa 1689
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgttaaa 240

Qy   1690 ttcattttattcttattaaggaaatactttgcataaattaatcagccccagagtgcttta 1749
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ttcattttattcttattaaggaaatactttgcataaattaatcagccccagagtgcttta 300

Qy   1750 aaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaatatgta 1809
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 aaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaatatgta 360

Qy   1810 tcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggacgtgca 1869
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 tcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggacgtgca 420

Qy   1870 gacaggctggcattctagattacttttcttactctgaaacatggcctgtttgggagtgcg 1929
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 gacaggctggcattctagattacttttcttactctgaaacatggcctgtttgggagtgcg 480

Qy   1930 ggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatcttatcttt 1989
          ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db    481 ggattcaaaggtggtcccacsgctgcccctactgcaaatggcagttttaatcttatcttt 540

Qy   1990 tggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatctctgt 2049
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy   1891 actttttcttactctgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccacc 1950
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15167 actttttcttactctgaaacatggcctgtttgggagtgcgggattcaaaggtggtcccacg 15226

Qy   1951 gctgcccctactgcaaatggcagttttaatcttatcttttggcttctgcagatggttgca 2010
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15227 gctgcccctactgcaaatggcagttttaatcttatcttttggcttctgcagatggttgca 15286

Qy   2011 attgatccttaaccaataatggtcagtcctcatctctgtcctgcttcataggtgccacct 2070
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  15287 attgatccttaaccaataatggtcagtcctcatctctgtcgtgcttcataggtgccacct 15346

Qy   2071 tgtgtgtttaaagaagggaagctttgtacctttagagtgtaggtgaaatgaatgaatggc 2130
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15347 tgtgtgtttaaagaagggaagctttgtacctttagagtgtaggtgaaatgaatgaatggc 15406

Qy   2131 ttggagtgcactgagaacagcatatgatttcttgctttggggaaaaagaatgatgctatg 2190
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15407 ttggagtgcactgagaacagcatatgatttcttgctttggggaaaaagaatgatgctatg 15466

Qy   2191 aaattggtgggtggtgtatttgagaagataatcattgcttatgtcaaatggagctgaatt 2250
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15467 aaattggtgggtggtgtatttgagaagataatcattgcttatgtcaaatggagctgaatt 15526

Qy   2251 tgataaaaacccaaaatacagctatgaagtgctgggcaagtttacttttttttctgatgtt 2310
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15527 tgataaaaacccaaaatacagctatgaagtgctgggcaagtttacttttttttctgatgtt 15586

Qy   2311 tcctacaact 2320
          ||||||||||
Db  15587 tcctacaact 15596
```

RESULT 7
AAH51403
ID AAH51403 standard; DNA; 1001 BP.
XX
AC AAH51403;
XX
DT 29-AUG-2001 (first entry)
XX
DE Human UGT1A7 related DNA containing a biallelic polymorphism SEQ ID 294.
XX
KW Human; biallelic marker; single nucleotide polymorphism; SNP; MGSTII;
KW microsomal glutathione S-transferase II; malate decarboxylase enzyme;
KW DME1; ME1; cytochrome P450; glutathione reductase; GSHR; GSHS; GGT5;
KW flavin-containing monooxygenase; FMO; gamma-glutamyltransferase 5;
KW dipeptidase; DP; glucose 6-phosphate dehydrogenase; G6PDH; haplotype;
KW phosphogluconate dehydrogenase; PGDH; drug metabolism; phenotype;
KW uridine diphosphate glucoronosyl transferase; UGT2; asthma; hepatoxicity;
KW zileuton; ds.
XX
OS Homo sapiens.
XX
PN WO200058508-A2.
XX
PD 05-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-IB00403.
XX
PR 25-MAR-1999; 99US-0126269.
PR 30-APR-1999; 99US-0131961.
XX
PA (GEST) GENSET.
XX
PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A;
XX
DR WPI; 2000-638353/61.
XX
PT Polynucleotides comprising sequences from malate decarboxylase
PT enzyme-related biallelic markers used for genotyping -
XX
PS Claim 13; Page 487; 673pp; English.

XX
CC Sequences AAH51110-AAH51593 represent human DNA fragments which contain
CC biallelic markers. The sequences are related to various human genes
CC including microsomal glutathione S-transferase II (MGSTII), malate
CC decarboxylase enzyme (DME1/ME1), cytochrome P450, glutathione
CC reductase/synthase (GSHR/GSHS), flavin-containing monooxygenases (FMO),
CC gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 6-phosphate
CC dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine
CC diphosphate glucoronosyl transferases (UGT2). Each of these sequences
CC contains a biallelic marker/polymorphism, which is represented in the
CC sequence as a degenerate/undefined base. The genes to which the biallelic
CC marker containing sequences are related are involved in drug metabolism.
CC Sequences AAH51594 - AAH51598 represent the genomic sequence of the
CC MGSTII gene and four alternative MGSTII cDNA sequences. AAB62905-AAB62906
CC are MGSTII gene products. PCR primers AAH51599 and AAH51600 are used in
CC an example for the amplification of human genomic DNA fragments. The
CC invention includes a method of genotyping comprising determining the
CC identity of a nucleotide at a DME- or MGSTII-related biallelic marker in
CC a biological sample. The method is used to determine the frequency in
CC population of an allele of a DME- or MGST-II related biallelic marker and
CC to select an individual for inclusion in a clinical trial of a drug
CC treatment. The method is also used to detect association between allele
CC and phenotype, and to detect association between haplotype and phenotype.
CC The polynucleotides are used, in hybridization assays, sequencing assays
CC or allele specific amplification assays. The method can be used to
CC determine whether an individual suffers or is at risk of developing
CC asthma or is at risk of developing hepatoxicity on treatment with
CC zileuton.
XX
SQ Sequence 1001 BP; 269 A; 205 C; 233 G; 293 T; 1 other;

Query Match 43.0%; Score 998.6; DB 21; Length 1001;
Best Local Similarity 99.9%; Pred. No. 2.1e-273;
Matches 998; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1322 agaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctgg 1381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 agaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctgg 60

Qy   1382 ccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcag 1441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcag 120

Qy   1442 cccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccg 1501
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 cccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccg 180

Qy   1502 tcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgct 1561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 tcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgct 240

Qy   1562 tggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtg 1621
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 tggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtg 300

Qy   1622 ggaaataaggtaaaatttgaaccattccctagtcatttccaaacttgaaaacagaatca 1681
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ggaaataaggtaaaatttgaaccattccctagtcatttccaaacttgaaaacagaatca 360

Qy   1682 gtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccaga 1741
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccaga 420

Qy   1742 gtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttg 1801
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 gtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttg 480

Qy   1802 aatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagag 1861
          ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db    481 aatatgtatcgtgccccctcyggtgtctttgatcaggatgacatgtgccatttttcagag 540

Qy   1862 gacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttg 1921
```

```
XX
FH   Key             Location/Qualifiers
FT   variation       replace (2510,C)
FT                   /*tag= a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   CDS             2616..14887
FT                   /*tag= b
FT                   /product= "UDP glycosyltransferase 1"
FT                   /note= "This CDS is specifically claimed in claim
FT                        26"
FT   variation       replace (2756,T)
FT                   /*tag= c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (2826,A)
FT                   /*tag= d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (3155,G)
FT                   /*tag= e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (9508,C)
FT                   /*tag= f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (9511,T)
FT                   /*tag= g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (10091,T)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (10094,T)
FT                   /*tag= i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (10095,C)
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (10140,C)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (14423,T)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (14713,T)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (14776,T)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (14971,C)
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   WO200179230-A2.
XX
PD   25-OCT-2001.
XX
PF   13-APR-2001; 2001WO-US12273.
XX
PR   18-APR-2000; 2000US-197514P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Chew A,  Choi JY,  Koshy B,  Rounds E;
XX
DR   WPI; 2002-075063/10.
DR   P-PSDB; AAU10940.
XX
PT   Genotyping a human UDP glycosyltransferase 1 gene of an individual for
PT   determining the haplotype of an individual, involves determining the
PT   identity of a nucleotide pair at specific polymorphic sites for two
PT   copies of the gene  -
XX
PS   Claim 21; Fig 1; 81pp; English.
XX
CC   The invention relates to genotyping a human UDP glycosyltransferase
CC   (UGT1A1) gene of an individual, involving determining for the two copies
CC   of the UGT1A1 gene present in the individual, the identity of the
CC   nucleotide pair at one or more polymorphic sites. The new method is
CC   useful for determining whether an individual has a haplotype or haplotype
CC   pairs, given in the specification. It is useful for improving the
CC   efficacy and reliability of several steps in the discovery and
CC   development of drugs for treating diseases associated with UGT1A1
CC   activity, e.g., Gilbert's syndrome and Crigler-Najjar syndrome, to
CC   validate UGT1A1 as a candidate agent for treating a specific condition or
CC   disease predicted to be associated with UGT1A1 activity, and in the
CC   design of clinical trials of candidate drugs for treating a specific
CC   condition or disease predicted to be associated with UGT1A1 activity. The
CC   method is useful to screen for compounds targeting UGT1A1 to treat a
CC   specific condition or disease associated with UGT1A1 activity. A nucleic
CC   acid (I) comprising a polymorphic variant of a reference sequence for
CC   the UGT1A1 gene or cDNA (II) or its fragment is useful in studying the
CC   expression and function of UGT1A1, and in expressing UGT1A1 protein for
CC   use in screening for candidate drugs to treat diseases related to UGT1A1
CC   activity. (I) or (II) is useful for therapeutic purposes. (II) or a
CC   recombinant organism comprising (II) is useful for studying
CC   expression of the UGT1A1 isogenes in vivo, for in vivo screening and
CC   testing of drugs targeted against UGT1A1 protein, and for testing the
CC   efficacy of therapeutic agents and compounds for Gilbert's syndrome and
CC   Crigler-Najjar syndrome, in a biological system. The present sequence
CC   represents the coding sequence of UDP glycosyltransferase 1.
XX
SQ   Sequence 18887 BP; 5079 A; 4037 C; 4131 G; 5640 T; 0 other;


  Query Match             43.3%;  Score 1004.4;  DB 24;  Length 18887;
  Best Local Similarity   98.4%;  Pred. No. 2.8e-274;
  Matches 1014;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy    1291 aaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctccagccttcac 1350
           || | |    ||  ||     ||||||||||||||||||||||||||||||||||||||||||
Db   14567 aacgaaactgtctttgtgtttagttacaaggagaacatcatgcgcctctccagccttcac 14626

Qy    1351 aaggaccgcccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgagg 1410
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14627 aaggaccgcccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgagg 14686

Qy    1411 cacaagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccat 1470
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14687 cacaagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccat 14746

Qy    1471 tccttggacgtgattggtttcctcttggccgtcgtgctgacagtggccttcatcaccttt 1530
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14747 tccttggacgtgattggtttcctcttggccgtcgtgctgacagtggccttcatcaccttt 14806

Qy    1531 aaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcgagttaagaaagcc 1590
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14807 aaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcgagttaagaaagcc 14866

Qy    1591 cacaaatccaagacccattgagaagtgggtgggaaataaggtaaaattttgaaccattcc 1650
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14867 cacaaatccaagacccattgagaagtgggtgggaaataaggtaaaattttgaaccattcc 14926

Qy    1651 ctagtcatttccaaacttgaaaacagaatcagtgttaaattcattttattcttattaagg 1710
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14927 ctagtcatttccaaacttgaaaacagaatcagtgttaaattcattttattcttattaagg 14986

Qy    1711 aaatactttgcataaattaatcagccccagagtgctttaaaaaattctcttaaataaaaa 1770
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14987 aaatactttgcataaattaatcagccccagagtgctttaaaaaattctcttaaataaaaa 15046

Qy    1771 taatagactcgctagtcagtaaagatatttgaatatgtatcgtgccccctccggtgtctt 1830
           ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   15047 taatagactcgctagtcagtaaagatatttgaatatgtatcgtgccccctctggtgtctt 15106

Qy    1831 tgatcaggatgacatgtgccatttttcagaggacgtgcagacaggctggcattctagatt 1890
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15107 tgatcaggatgacatgtgccatttttcagaggacgtgcagacaggctggcattctagatt 15166
```

```
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1655 BP; 402 A; 388 C; 403 G; 462 T; 0 other;

  Query Match             66.9%;  Score 1551.8;  DB 23;  Length 1655;
  Best Local Similarity   97.7%;  Pred. No. 0;
  Matches 1574:  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy      1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccctcctctatgtgtg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     45 ctcagctgcagttctctgatggcttgcacagggtggaccagccccctcctctatgtgtg 104

Qy     61 tgtcagctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    105 tgtcagctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 164

Qy    121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggggcatgag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    165 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggggcatgag 224

Qy    181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    225 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 284

Qy    241 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    285 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 344

Qy    301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360
          |||| |||||||||||||||||||||||| | | |||||||||  || ||||||  ||||
Db    345 catgctcaatggaaagcacaagtacgaagtttgttttctctatttctgagttcatccaat 404

Qy    361 gacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagta 420
          |    ||||  ||||||||||||  ||||||||||||||||||||| |||  ||||||||||
Db    405 ggtttttttaacttatttttttcgcattgcaggagtttgtttaatgaccgaaaattagta 464

Qy    421 gaatacttaaaggagagttctttgatgcagtgtttctcgatccttttgataactgtggc 480
          |||||||||||||||||||||||||||||| |||||||| ||||||||||||  |||||
Db    465 gaatacttaaaggagagttctttgatgcggtgtttcttgatccttttgatgcctgtgcg 524

Qy    481 ttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgc 540
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||  ||||
Db    525 ttaattgttgccaaatatttctccctcccctctgtggtcttcgccaggggaataggttgc 584

Qy    541 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    585 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 644

Qy    601 ttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    645 ttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttg 704

Qy    661 gaggaacatttattatgccaccgttttttcaaaaatgccctagaaatagcctctgaaatt 720
          |||||||||||||| |||||   ||||| ||||||||||||||||||||||||||||||
Db    705 gaggaacatttatttgccagtatttttccaaaaatgccctagaaatagcctctgaaatt 764

Qy    721 ctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcga 780
          |||||||||||||| || |  ||||||||||||||||||||||||||||||||||||||
Db    765 ctccaaacacctgtcacagcatatgatctctacagccacacatcaatttggttgttgcga 824

Qy    781 acggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggt 840
          || |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    825 acagactttgttttggactatcccaaacccgtgatgcccaatatgatcttcattggtggt 884

Qy    841 atcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttct 900
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    885 atcaactgccatcagggaaagccattgcctatggaatttgaagcctacattaatgcttct 944

Qy    901 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    945 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 1004

Qy    961 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1005 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1064

Qy   1021 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1065 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaa 1124

Qy   1081 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1125 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1184

Qy   1141 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1185 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1244

Qy   1201 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1245 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1304

Qy   1261 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1305 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1364

Qy   1321 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1365 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1424

Qy   1381 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1425 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1484

Qy   1441 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1485 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1544

Qy   1501 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1545 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1604

Qy   1561 ttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattga 1611
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1605 ttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattga 1655


RESULT   6
AAS18543
ID   AAS18543 standard; DNA; 18887 BP.
XX
AC   AAS18543;
XX
DT   12-MAR-2002  (first entry)
XX
DE   DNA encoding UDP glycosyltransferase 1 (UGT1A1).
XX
KW   UDP glycosyltransferase 1; UGT1A1; human; haplotyping;
KW   drug discovery; Gilbert's syndrome; Crigler-Najjar syndrome; ds.
XX
OS   Homo sapiens.
```

```
Qy    858 aaagccgttgcctatggaatttgaagcctacattaatgcttctggagaacatggaattgt 917
Db    881 ggaagccactatctaggaatttgaagcctacattaatgcttctggagaacatggaattgt 940

Qy    918 ggttttctctttgggatcaatggtctcagaaattccagagaagaaagctatggcaattgc 977
Db    941 ggttttctctttggaatcaatggtctcagaaattccagagaagaaagctatggcaattgc 1000

Qy    978 tgatgctttgggcaaaatccctcagacagtcctgtggcggtacactggaacccgaccatc 1037
Db   1001 tgatgctttgggcaaaatccctcagacagtcctgtggcggtacactggaacccgaccatc 1060

Qy   1038 gaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacgatctgcttggtca 1097
Db   1061 gaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacgatctgcttggtca 1120

Qy   1098 cccgatgacccgtgcctttatcacccatgctggttcccatggtgtttatgaaagcatatg 1157
Db   1121 cccgatgacccgtgcctttatcacccatgctggttcccatggtgtttatgaaagcatatg 1180

Qy   1158 caatggcgttcccatggtgatgatgcccttgtttggtgatcagatggacaatgcaaagcg 1217
Db   1181 caatggcgttcccatggtgatgatgcccttgtttggtgatcagatggacaatgcaaagcg 1240

Qy   1218 catggagactaagggagctggagtgaccctgaatgttctggaaatgacttctgaagattt 1277
Db   1241 catggagactaagggagctggagtgaccctgaatgttctggaaatgacttctgaagattt 1300

Qy   1278 agaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcct 1337
Db   1301 agaaaatgctcaaaaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcct 1360

Qy   1338 ctccagccttcacaaggaccgcccggtggagccgctggacctggccgtgttctgggtgga 1397
Db   1361 ctccagccttcacaaggaccgcccggtggagccgctggacctggccgtgttctgggtgga 1420

Qy   1398 gtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagcccacgacctcacctg 1457
Db   1421 gtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagcccacgacctcacctg 1480

Qy   1458 gtaccagtaccattccttggacgtgattggtttcctcttggccgtcgtgctgacagtggc 1517
Db   1481 gtaccagtaccattccttggacgtgattggtttcctcttggccgtgctgctgacagtggc 1540

Qy   1518 cttcatcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcg 1577
Db   1541 cttcatcacctttaaatgttgtgcttatggctaccggaaatgcttggggaaaaaagggcg 1600

Qy   1578 agttaagaaagcccacaaatccaagacccattgagaagtgggtgggaaataaggtaaaat 1637
Db   1601 agttaagaaagcccacaaatccaagacccattgagaagtgggtgggaaataaggtaaaat 1660

Qy   1638 tttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgttaaattcatttt 1697
Db   1661 tttgaaccattccctagtcatttccaaacttgaaaacagaatcagtgttaaattcatttt 1720

Qy   1698 attcttattaaggaaatactttgcataaattaatcagccccagagtgctttaaaaaattc 1757
Db   1721 attcttattaaggaaatactttgcat-aattaatcagccccagagtgctttaaaaaattc 1779

Qy   1758 tcttaaataaaaataatagactcgctagtcagtaaagatatttgaatatgtatcgtgccc 1817
Db   1780 tcttaaataaaaataatagactcgctagtcagtaaagatatttgaatatgtatcgtgccc 1839

Qy   1818 cctccggtgtctttgatcaggatgacatgtgccatttttcagaggacgtgcagacaggct 1877
Db   1840 cctccggtgtctttgatcaggatgacatgtgccatttttcagaggacgtgcagacaggct 1899

Qy   1878 ggcattctagattacttttcttactctgaaacatggcctgtttgggagtgcgggattcaa 1937
Db   1900 ggcattctagattacttttcttactctgaaacatggcctgtttgggagtgcgggattcaa 1959

Qy   1938 aggtggtcccaccgctgcccctactgcaaatggcagttttaatcttatcttttggcttct 1997
Db   1960 aggtggtcccaccgctgcccctactgcaaatggcagttttaatcttatcttttggcttct 2019

Qy   1998 gcagatggttgcaattgatccttaaccaataatggtcagtcctcatctctgtcctgcttc 2057
Db   2020 gcagatggttgcaattgatccttaaccaataatggtcagtcctcatctctgtcctgcttc 2079

Qy   2058 ataggtgccacc-ttgtgtgtttaaagaagggaagctttgtacctttagagtgtaggtga 2116
Db   2080 ataggtgccacctttgtgtgtttaaagaagggaagctttgtacctttagagtgtaggtga 2139

Qy   2117 aatgaatgaatggcttggagtgcactgagaacagcatatgatttcttgctttggggaaaa 2176
Db   2140 aatgaatgaatggcttggagtgcactgagaacagcatatgatttcttgctttggggaaaa 2199

Qy   2177 agaatgatgctatgaaattggtgggtggtgtatttgagaagataatcattgcttatgtca 2236
Db   2200 agaatgatgctatgaaattggtgggtggtgtatttgagaagataatcattgcttatgtca 2259

Qy   2237 aatggagctgaatttgataaaaacccaaaatacagctatgaagtgctgggcaagtttact 2296
Db   2260 aatggagctgaatttgataaaaacccaaaatacagctatgaagtgctgggcaagtttact 2319

Qy   2297 ttttttctgatgtttcctacaact 2320
Db   2320 ttttttctgatgtttcctacaact 2343
```

RESULT 5
AAS91252
ID AAS91252 standard; cDNA; 1655 BP.
XX
AC AAS91252;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #27056.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG27065.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 27056; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques

```
Db   1870 gtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttggga 1929

Qy   1925 gtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatctta 1984
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1930 gtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaatctta 1989

Qy   1985 tcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatc 2044
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1990 tcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcctcatc 2049

Qy   2045 tctgtcctgcttcataggtgccacc-ttgtgtgtttaaagaagggaagctttgtaccttt 2103
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   2050 tctgtcctgcttcataggtgccacctttgtgtgtttaaagaagggaagctttgtaccttt 2109

Qy   2104 agagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttctt 2163
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2110 agagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgatttctt 2169

Qy   2164 gctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatc 2223
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2170 gctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagataatc 2229

Qy   2224 attgcttatgtcaaatggagctgaatttgataaaaaccccaaaatacagctatgaagtgct 2283
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2230 attgcttatgtcaaatggagctgaatttgataaaaaccccaaaatacagctatgaagtgct 2289

Qy   2284 gggcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||||
Db   2290 gggcaagtttacttttttttctgatgtttcctacaact 2326
```

RESULT 4
AAQ27370
ID AAQ27370 standard; cDNA; 2368 BP.
XX
AC AAQ27370;
XX
DT 27-JAN-1993 (first entry)
XX
DE HUG-Br2.
XX
KW Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
KW monoglucuronide; diglucuronide; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 30..801
FT /*tag= a
FT polyA_signal 2347..2352
FT /*tag= b
FT /number= 1
FT polyA_signal 2355..2360
FT /*tag= c
FT /number= 2
XX
PN WO9212987-A.
XX
PD 06-AUG-1992.
XX
PF 10-JAN-1992; 92WO-US00282.
XX
PR 10-JAN-1991; 91US-0639453.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
XX
PI Owens IS, Ritter JK;
XX
DR WPI; 1992-284593/34.
DR P-PSDB; AAR26154.
XX
PT Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT types I and II
XX
PS Disclosure; Fig 9A-I; 99pp; English.
XX
CC Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
CC been isolated. They are referred to as HUGBr1 (AAQ27369) and HUGBr2
CC (AAQ27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
CC upon expression individually in COS-1 cells, encode isoforms that
CC catalyse the formation of the two bilirubin monoglucuronides and
CC the diglucuronide.
CC The cDNAs contain identical 3' ends (1469 bp in length) to each
CC other and to that of the human phenol transferase cDNA, HLUGP1
CC (Harding et al., Proc. Natl. Aca. Sci. USA 85:8281 (1988)).
CC In contrast, they have unique 5' ends.
XX
SQ Sequence 2368 BP; 609 A; 532 C; 566 G; 661 T; 0 other;

Query Match 68.0%; Score 1576.8; DB 13; Length 2368;
Best Local Similarity 82.3%; Pred. No. 0;
Matches 1847; Conservative 0; Mismatches 392; Indels 5; Gaps 3;

```
Qy    81 ctttgccgaggcagggaagctactggtagtgcccatggatgggagccactggttcaccat 140
         ||  || |||   || ||| |  |||| |||||||  ||||| |||| |||| ||| |||
Db   101 ctgggctgagagtggaaaggtgttggtggtgcccactgatggcagcccctggctcagcat 160

Qy   141 gaggtcggtggtggagaaactcattctcaggggggcatgaggtggttgtagtcatgccaga 200
         | ||  ||  ||  | | |||  |  ||| || ||  ||| ||| ||  |||  |||
Db   161 gcgggaggccttgcgggagctccatgccagaggccaccaggcggtggtcctcaccccaag 220

Qy   201 ggtgagttggcaactgggaagatcactgaattgcacagtgaagacttattcaacttcata 260
         |    |   ||    |  |        || |||  |||   | ||| |   | |||
Db   221 aggtgaatatgcacatcaagaagagaaatttttcaccctgacagcctatgctgttccatg 280

Qy   261 taccctggaggatctggaccgggagttcaaggcttttgcccatgctcaat---ggaaagc 317
         |||| | ||||  | || || |       ||  |   | || |     |     | ||
Db   281 gacccagaaggaatttgatcgcgttacgctgggctacactcaagggttctttgaaacaga 340

Qy   318 acaagtacgaagtatatattctctattaatgggttcatacaatgacatttttgacttatt 377
         |||  | |  |  | |||||||  |   ||||        ||  |  | | |       |
Db   341 acatcttctgaagagatattctagaagtatggcaattatgaacaatgtatctttggccct 400

Qy   378 tttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaatacttaaaggagag 437
         |  |     |||  |    |  |  |  |  |     |  |     || | || |  |
Db   401 tcataggtgttgtgtggagctactgcataatgaggccctgatcaggcacctgaatgctac 460

Qy   438 ttcttttgatgcagtgtttctcgatccttttgataactgtggcttaattgttgccaaata 497
         ||| |||||||  || ||    || ||  || |   ||| ||      |  | || || ||
Db   461 ttcctttgatgtggttttaacagaccccgttaacctctgcggggcggtgctggctaagta 520

Qy   498 tttctccctcccctccgtggtcttcgccaggggaatactttgccactatcttgaagaagg 557
           | ||  | ||  | ||| | ||    |||  || |  ||  |||         | ||
Db   521 cctgtcgattcctgctgtgttttttggaggtacattccatgtgacttagactttaaggg 580

Qy   558 tgcacagtgccctgctcctctttcctatgtccccagaattctcttagggttctcagatgc 617
           |||||||  ||   ||||   |||||| | || |   | ||          |||||
Db   581 cacacagtgtccaaatccttcctcctatattcctaagttactaacgaccaattcagacca 640

Qy   618 catgactttcaaggagagagtacggaaccacatcatgcacttggaggaacatttattatg 677
         ||||||  |||  | | || ||   ||||    ||   |  |||         |  | ||
Db   641 catgacattcctgcaaagggtcaagaacatgctctaccctctggccctgtcctacatttg 700

Qy   678 ccaccgtttcttcaaaaatgccctagaaatagcctctgaaattctccaaacacctgttac 737
         |||  |||||       |      |    | ||||||||  || | || | |   || ||
Db   701 ccatactttttctgccccttatgcaagtcttgcctctgagctttttcagagagaggtgac 760

Qy   738 ggagtatgatctctacagccacacatcaatttggttgttgcgaacggactttgttttgga 797
          |   | || || |  |||    ||||  | ||| ||||  |||  ||||||||   |||
Db   761 tgtccaggacctattgagctctgcatctgtctggctgtttagaagtgactttgtgaagga 820

Qy   798 ctatcccaaacccgtgatgcccaacatgatcttcattggtggtatcaactgccatcaggg 857
          || || |  ||| | |||||||| ||| |||||||||| || ||||||||    ||  |
Db   821 ttaccctaggcccatcatgcccaatatggtcttcattgggggcatcaactgttgccaacg 880
```

```
PT   Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT   for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT   types I and II
XX
PS   Disclosure; Fig 9A-I; 99pp; English.
XX
CC   Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
CC   been isolated.  They are referred to as HUGBr1 (AAQ27369) and HUGBr2
CC   (AAQ27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
CC   upon expression individually in COS-1 cells, encode isoforms that
CC   catalyse the formation of the two bilirubin monoglucuronides and
CC   the diglucuronide.
CC   The cDNAs contain identical 3' ends (1469 bp in length) to each
CC   other and to that of the human phenol transferase cDNA, HLUGP1
CC   (Harding et al., Proc. Natl. Aca. Sci. USA 85:8281 (1988)).
CC   In contrast, they have unique 5' ends.
XX
SQ   Sequence 2351 BP; 602 A; 540 C; 556 G; 653 T; 0 other;


  Query Match             68.2%;  Score 1581.4;  DB 13;  Length 2351;
  Best Local Similarity   82.5%;  Pred. No. 0;
  Matches 1861;  Conservative    0;  Mismatches  391;  Indels    5;  Gaps    4;

Qy      68 tgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggatgggagcc 127
           |||||  |   |   |  || | || ||||||| ||||| |  | ||  |||||||| ||||
Db      71 tgctgggcccagtggtgtcccatgctgggaagatactgttgatcccagtggatggcagcc 130

Qy     128 actggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgaggtggttg 187
           ||||| | | ||||     ||   |  || | ||    |  ||||| |||||  | ||||
Db     131 actggctgagcatgcttggggccatccagcagctgcagcagaggggacatgaaatagttg 190

Qy     188 tagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtgaagactt 247
           |  |    || || |     | | |  |  ||     |     || |||  ||||||| |
Db     191 tcctagcacctgacgcctcgttgtacatcagagacggagcattttacaccttgaagacgt 250

Qy     248 attcaacttcatataccctggaggatctggaccgggagtt--caaggcttttgcccatgc 305
           |  |     |||       |||||||  || |   |   ||    || ||   ||  |
Db     251 accctgtgccattccaaagggaggatgtgaaagagtcttttgttagtctcgggcataatg 310

Qy     306 tcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaatgacat 365
           |   ||  ||  |    | |        | |  || |   ||      | | ||  |
Db     311 tttttgagaatgattctttcctgcagcgtgtgatcaaaacatacaagaaaataaaaaagg 370

Qy     366 ttttgact-tatttttttcaaattgcaggagtttgtttaaagacaaaaaattagtagaat 424
             |   || |  |||| ||    ||       ||  |  |  ||||  |  |  | |  |
Db     371 actctgctatgcttttgtctggctgttcccacttactgcacaacaaggagctcatggcct 430

Qy     425 acttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggcttaa 484
            | |    || ||    |||||||   || |    || |||||   |   ||  ||   |
Db     431 ccctggcagaaagcagctttgatgtcatgctgacggacccttttccttccttgcagcccca 490

Qy     485 ttgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgccact 544
           | || ||| | ||  | || || ||| | ||  |||||      | | | |  |||  |
Db     491 tcgtggcccagtacctgtctctgcccactgtattcttcttgcatgcactgccatgcagcc 550

Qy     545 atcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctcttag 604
                || | | |||||||||    ||  | ||||| || |||||   |||||
Db     551 tggaatttgaggctacccagtgccccaacccattctcctacgtgcccaggcctctctcct 610

Qy     605 ggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttggagg 664
               ||||||  |||||| |||  | || | ||   ||||    ||||   |||
Db     611 ctcattcagatcacatgaccttcctgcagcgggtgaagaacatgctcattgccttttcac 670

Qy     665 aacatttattatgccaccgtttttttcaaaaatgccctagaaatagcctctgaaattctcc 724
           |  | ||  | ||| ||    |||              |    | ||||| ||| | || |
Db     671 agaactttctgtgcgacgtggtttattccccgtatgcaacccttgcctcagaattccttc 730

Qy     725 aaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcgaacgg 784
           | | |   || || |   | || || |  |||    ||||  | ||| ||||  |||  |
Db     731 agagagaggtgactgtccaggacctattgagctctgcatctgtctggctgtttagaagtg 790

Qy     785 actttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggtatca 844
```

```
                ||||||   ||| || || |  ||| | |||||||| ||| | ||  ||||||| ||||
Db     791 actttgtgaaggattaccctaggcccatcatgcccaatatggtttttgttggtggaatca 850

Qy     845 actgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttctggag 904
           |||||| |||   ||| ||  |  |  ||||||||||||||||||||||||||||||||
Db     851 actgccttcaccaaaatccactatcccaggaatttgaagcctacattaatgcttctggag 910

Qy     905 aacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaagaaag 964
           ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     911 aacatggaattgtggttttctctttggaatcaatggtctcagaaattccagagaagaaag 970

Qy     965 ctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactg 1024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     971 ctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactg 1030

Qy    1025 gaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctacccaaaacg 1084
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1031 gaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctacccaaaacg 1090

Qy    1085 atctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgttt 1144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1091 atctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgttt 1150

Qy    1145 atgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatgg 1204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1151 atgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatgg 1210

Qy    1205 acaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatga 1264
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1211 acaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatga 1270

Qy    1265 cttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaaggaga 1324
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    1271 cttctgaagatttagaaaatgctcaaaaagcagtcatcaatgacaaaagttacaaggaga 1330

Qy    1325 acatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccg 1384
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1331 acatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccg 1390

Qy    1385 tgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagccc 1444
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1391 tgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagccc 1450

Qy    1445 acgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtcg 1504
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1451 acgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtgc 1510

Qy    1505 tgctgacagtggccttcatcaccttaaatgttgtgcttatggctaccggaaatgcttgg 1564
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1511 tgctgacagtggccttcatcaccttaaatgttgtgcttatggctaccggaaatgcttgg 1570

Qy    1565 ggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtggga 1624
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1571 ggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggtggga 1630

Qy    1625 aataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtg 1684
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1631 aataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatcagtg 1690

Qy    1685 ttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccagagtg 1744
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1691 ttaaattcattttattcttattaaggaaatactttgcat-aattaatcagccccagagtg 1749

Qy    1745 ctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaat 1804
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1750 ctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatatttgaat 1809

Qy    1805 atgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggac 1864
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1810 atgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcagaggac 1869

Qy    1865 gtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgtttggga 1924
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy    781 acggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggt 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 acggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggt 840

Qy    841 atcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttct 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 atcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttct 900

Qy    901 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 960

Qy    961 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1020

Qy   1021 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctacccaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctacccaa 1080

Qy   1081 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1140

Qy   1141 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1200

Qy   1201 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1260

Qy   1261 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1320

Qy   1321 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1380
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1380

Qy   1381 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1440
          ||||.||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1440

Qy   1441 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1500

Qy   1501 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1560

Qy   1561 ttggcgaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggt 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggt 1620

Qy   1621 gggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 gggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatc 1680

Qy   1681 agtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccag 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 agtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccag 1740

Qy   1741 agtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatattt 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 agtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatattt 1800

Qy   1801 gaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcaga 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 gaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcaga 1860
```

```
Qy   1861 ggacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgttt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 ggacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgttt 1920

Qy   1921 gggagtgccggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaat 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 gggagtgccggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaat 1980

Qy   1981 cttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcct 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 cttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcct 2040

Qy   2041 catctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacc 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 catctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacc 2100

Qy   2101 tttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgattt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 tttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgattt 2160

Qy   2161 cttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagata 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 cttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagata 2220

Qy   2221 atcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagt 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 atcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagt 2280

Qy   2281 gctgggcaagtttacttttttttctgatgtttcctacaact 2320
          ||||||||||||||||||||||||||||||||||||||||
Db   2281 gctgggcaagtttacttttttttctgatgtttcctacaact 2320
```

```
RESULT   3
AAQ27369
ID   AAQ27369 standard; cDNA; 2351 BP.
XX
AC   AAQ27369;
XX
DT   27-JAN-1993  (first entry)
XX
DE   HUG-Br1.
XX
KW   Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
KW   monoglucuronide; diglucuronide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             16..784
FT                   /*tag= a
FT   polyA_signal    2330..2335
FT                   /*tag= b
FT                   /number= 1
FT   polyA_signal    2338..2343
FT                   /*tag= c
FT                   /number= 2
XX
PN   WO9212987-A.
XX
PD   06-AUG-1992.
XX
PF   10-JAN-1992;   92WO-US00282.
XX
PR   10-JAN-1991;   91US-0639453.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Owens IS,  Ritter JK;
XX
DR   WPI; 1992-284593/34.
DR   P-PSDB; AAR26153.
XX
```

```
Qy   1801 gaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcaga 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 gaatatgtatcgtgccccctccggtgtctttgatcaggatgacatgtgccatttttcaga 1860

Qy   1861 ggacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgttt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 ggacgtgcagacaggctggcattctagattacttttcttactctgaaacatggcctgttt 1920

Qy   1921 gggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaat 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 gggagtgcgggattcaaaggtggtcccaccgctgcccctactgcaaatggcagttttaat 1980

Qy   1981 cttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcct 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 cttatcttttggcttctgcagatggttgcaattgatccttaaccaataatggtcagtcct 2040

Qy   2041 catctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacc 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 catctctgtcctgcttcataggtgccaccttgtgtgtttaaagaagggaagctttgtacc 2100

Qy   2101 tttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgattt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 tttagagtgtaggtgaaatgaatgaatggcttggagtgcactgagaacagcatatgattt 2160

Qy   2161 cttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagata 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 cttgctttggggaaaaagaatgatgctatgaaattggtgggtggtgtatttgagaagata 2220

Qy   2221 atcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagt 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 atcattgcttatgtcaaatggagctgaatttgataaaaacccaaaatacagctatgaagt 2280

Qy   2281 gctgggcaagtttacttttttttctgatgtttcctacaact 2320
          |||||||||||||||||||||||||||||||||||||||
Db   2281 gctgggcaagtttacttttttttctgatgtttcctacaact 2320
```

RESULT 2
ABA05199
ID ABA05199 standard; cDNA; 2320 BP.
XX
AC ABA05199;
XX
DT 04-MAR-2002 (first entry)
XX
DE Human uridine 5'diphosphate glucuronyltransferase UGT1A9 cDNA #2.
XX
KW Human; uridine 5'diphosphate glucuronyltransferase enzyme; UGT1A9;
KW flavopiridol; cancer; gastrointestinal disease; parasitic infection;
KW cytostatic; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 19..1611
FT /*tag= a
FT /product= "UGT1A9"
XX
PN WO200180896-A2.
XX
PD 01-NOV-2001.
XX
PF 12-APR-2001; 2001WO-US12526.
XX
PR 21-APR-2000; 2000US-0553829.
XX
PA (ARCH-) ARCH DEV CORP.
XX
PI Ratain MJ, Innocenti F, Iyer L;
XX
DR WPI; 2002-075093/10.
DR P-PSDB; ABB04430.
XX
PT Combinations of flavopiridol and an agent that increases conjugative
PT enzyme activity or glucuronosyltransferase activity, with reduced side
PT effects, for treating cancer -
XX
PS Claim 81; Page 141-144; 145pp; English.
XX
CC The present invention relates to a method of reducing the toxicity of
CC flavopiridol by administration in combination with a second agent that
CC increases conjugative enzyme activity or glucuronosyltransferase
CC activity. This second agent should be capable of inhibiting biliary
CC transport and may be a uridine 5'diphosphate glucuronyltransferase such
CC as that encoded by the present sequence (designated UGT1A9). The method
CC can be used in the treatment of cancer, gastrointestinal diseases and
CC parasitic diseases.
XX
SQ Sequence 2320 BP; 608 A; 495 C; 547 G; 670 T; 0 other;

Query Match 100.0%; Score 2320; DB 24; Length 2320;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60

Qy     61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120

Qy    121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgag 180

Qy    181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240

Qy    241 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 300

Qy    301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360

Qy    361 gacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagta 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagta 420

Qy    421 gaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 gaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggc 480

Qy    481 ttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgc 540

Qy    541 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 600

Qy    601 ttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ttagggttctcagatgccatgactttcaaggagagagtacggaaccacatcatgcacttg 660

Qy    661 gaggaacatttattatgccaccgtttttttcaaaaatgccctagaaatagcctctgaaatt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 gaggaacatttattatgccaccgtttttttcaaaaatgccctagaaatagcctctgaaatt 720

Qy    721 ctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcga 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcga 780
```

XX
DR WPI; 2002-075093/10.
DR P-PSDB; ABB04429.
XX
PT Combinations of flavopiridol and an agent that increases conjugative
PT enzyme activity or glucuronosyltransferase activity, with reduced side
PT effects, for treating cancer -
XX
PS Claim 80; Page 137-140; 145pp; English.
XX
CC The present invention relates to a method of reducing the toxicity of
CC flavopiridol by administration in combination with a second agent that
CC increases conjugative enzyme activity or glucuronosyltransferase
CC activity. This second agent should be capable of inhibiting biliary
CC transport and may be a uridine 5'diphosphate glucuronyltransferase such
CC as that encoded by the present sequence (designated UGT1A9). The method
CC can be used in the treatment of cancer, gastrointestinal diseases and
CC parasitic diseases.
XX
SQ Sequence 2320 BP; 608 A; 495 C; 547 G; 670 T; 0 other;

Query Match 100.0%; Score 2320; DB 24; Length 2320;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ctcagctgcagttctctgatggcttgcacagggtggaccagccccttcctctatgtgtg 60

Qy     61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 tgtctgctgctgacctgtggctttgccgaggcagggaagctactggtagtgcccatggat 120

Qy    121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gggagccactggttcaccatgaggtcggtggtggagaaactcattctcaggggcatgag 180

Qy    181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 gtggttgtagtcatgccagaggtgagttggcaactgggaagatcactgaattgcacagtg 240

Qy    241 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 aagacttattcaacttcatataccctggaggatctggaccgggagttcaaggcttttgcc 300

Qy    301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 catgctcaatggaaagcacaagtacgaagtatatattctctattaatgggttcatacaat 360

Qy    361 gacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagta 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gacatttttgacttatttttttcaaattgcaggagtttgtttaaagacaaaaaattagta 420

Qy    421 gaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 gaatacttaaaggagagttcttttgatgcagtgtttctcgatccttttgataactgtggc 480

Qy    481 ttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ttaattgttgccaaatatttctccctcccctccgtggtcttcgccaggggaatactttgc 540

Qy    541 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 cactatcttgaagaaggtgcacagtgccctgctcctctttcctatgtccccagaattctc 600

Qy    601 ttagggttctcagatgccatgacttttcaaggagagagtacggaaccacatcatgcacttg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ttagggttctcagatgccatgacttttcaaggagagagtacggaaccacatcatgcacttg 660

Qy    661 gaggaacatttattatgccaccgtttttttcaaaaatgccctagaaatagcctctgaaatt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 gaggaacatttattatgccaccgtttttttcaaaaatgccctagaaatagcctctgaaatt 720

Qy    721 ctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcga 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ctccaaacacctgttacggagtatgatctctacagccacacatcaatttggttgttgcga 780

Qy    781 acggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggt 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 acggactttgttttggactatcccaaacccgtgatgcccaacatgatcttcattggtggt 840

Qy    841 atcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttct 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 atcaactgccatcagggaaagccgttgcctatggaatttgaagcctacattaatgcttct 900

Qy    901 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ggagaacatggaattgtggttttctctttgggatcaatggtctcagaaattccagagaag 960

Qy    961 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 aaagctatggcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtac 1020

Qy   1021 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 actggaacccgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaa 1080

Qy   1081 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 aacgatctgcttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggt 1140

Qy   1141 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 gtttatgaaagcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcag 1200

Qy   1201 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 atggacaatgcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaa 1260

Qy   1261 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 atgacttctgaagatttagaaaatgctctaaaagcagtcatcaatgacaaaagttacaag 1320

Qy   1321 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 gagaacatcatgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctg 1380

Qy   1381 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 gccgtgttctgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgca 1440

Qy   1441 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 gcccacgacctcacctggtaccagtaccattccttggacgtgattggtttcctcttggcc 1500

Qy   1501 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 gtcgtgctgacagtggccttcatcacctttaaatgttgtgcttatggctaccggaaatgc 1560

Qy   1561 ttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggt 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ttggggaaaaaagggcgagttaagaaagcccacaaatccaagacccattgagaagtgggt 1620

Qy   1621 gggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 gggaaataaggtaaaattttgaaccattccctagtcatttccaaacttgaaaacagaatc 1680

Qy   1681 agtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccag 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 agtgttaaattcattttattcttattaaggaaatactttgcataaattaatcagccccag 1740

Qy   1741 agtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatattt 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 agtgctttaaaaaattctcttaaataaaaataatagactcgctagtcagtaaagatattt 1800
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 21, 2002, 12:45:22 ; Search time 682.45 Seconds
(without alignments)
5836.679 Million cell updates/sec

Title: US-09-835-082-3
Perfect score: 2320
Sequence: 1 ctcagctgcagttctctgat.....ttctgatgtttcctacaact 2320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2320	100.0	2320	24	ABA05198	Human uridine 5'di
2	2320	100.0	2320	24	ABA05199	Human uridine 5'di
3	1581.4	68.2	2351	13	AAQ27369	HUG-Br1. Homo sap
4	1576.8	68.0	2368	13	AAQ27370	HUG-Br2. Homo sap
5	1551.8	66.9	1655	23	AAS91252	DNA encoding novel
6	1004.4	43.3	18887	24	AAS18543	DNA encoding UDP g
7	998.6	43.0	1001	21	AAH51403	Human UGT1A7 relat
8	870.6	37.5	1001	21	AAH51404	Human UGT1A7 relat
9	798	34.4	1667	13	AAQ33027	Exon 5 from the UG
10	769.6	33.2	1001	21	AAH51405	Human UGT1A7 relat
11	766.6	33.0	930	21	AAZ45116	UDP-glucuronosyltr
12	741.2	31.9	951	21	AAZ45115	UDP-glucuronosyltr
13	735	31.7	735	21	AAZ45118	UDP-glucuronosyltr
14	659.8	28.4	759	21	AAZ45117	UDP-glucuronosyltr
15	578.8	24.9	874	24	ABL01468	Murine apoptosis r
16	364.2	15.7	391	20	AAZ07208	Human lung tumour
17	364.2	15.7	391	21	AAC79066	Human lung tumour-
18	364.2	15.7	391	23	AAD23141	Human lung tumour-
19	359	15.5	1001	21	AAH51406	Human UGT1A7 relat
20	353	15.2	1001	21	AAH51407	Human UGT1A7 relat
21	345.2	14.9	1854	21	AAZ95200	Human UDP-glucuron
22	345	14.9	1766	23	AAS91380	DNA encoding novel
23	344.4	14.8	2107	19	AAV15900	Uridine diphospho-
24	342	14.7	380	21	AAC98723	Human colon cancer
25	339.8	14.6	2092	21	AAZ95199	Human UDP-glucuron
26	333.2	14.4	1976	21	AAZ95206	Human UDP-glucuron
27	332.8	14.3	1650	21	AAC65396	Human carbohydrate
28	327.2	14.1	2802	23	AAS69711	DNA encoding novel
29	325.6	14.0	2111	23	AAS69712	DNA encoding novel
30	322.4	13.9	1636	24	AAD24666	Human drug metabol
31	322.4	13.9	1722	22	AAD06860	Human breast cance
32	322.4	13.9	2974	22	AAS46185	Human DNA encoding
33	318	13.7	1859	23	AAS69710	DNA encoding novel
34	314.6	13.6	1713	22	AAD06844	Human breast cance
35	302.4	13.0	508	21	AAC98478	Human colon cancer
36	296	12.8	1001	21	AAH51408	Human UGT1A7 relat
37	282	12.2	2025	23	AAS91253	DNA encoding novel
38	230.4	9.9	366	21	AAA87085	Rat hepatocyte car
39	222.4	9.6	775	13	AAQ33026	Exons 3-4 from the
40	220.8	9.5	273	21	AAA87491	Rat hepatocyte car
c 41	203	8.8	2448	23	AAS72022	DNA encoding novel
c 42	203	8.8	2448	23	AAS74393	DNA encoding novel
c 43	203	8.8	2448	23	AAS78611	DNA encoding novel
c 44	203	8.8	2448	23	AAS78773	DNA encoding novel
c 45	203	8.8	2448	23	AAS84057	DNA encoding novel

ALIGNMENTS

RESULT 1
ABA05198
ID ABA05198 standard; cDNA; 2320 BP.
XX
AC ABA05198;
XX
DT 04-MAR-2002 (first entry)
XX
DE Human uridine 5'diphosphate glucuronyltransferase UGT1A9 cDNA #1.
XX
KW Human; uridine 5'diphosphate glucuronyltransferase enzyme; UGT1A9;
KW flavopiridol; cancer; gastrointestinal disease; parasitic infection;
KW cytostatic; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 19..1611
FT /*tag= a
FT /product= "UGT1A9"
XX
PN WO200180896-A2.
XX
PD 01-NOV-2001.
XX
PF 12-APR-2001; 2001WO-US12526.
XX
PR 21-APR-2000; 2000US-0553829.
XX
PA (ARCH-) ARCH DEV CORP.
XX
PI Ratain MJ, Innocenti F, Iyer L;